Transregio 391

TRR 391 Working Paper #3

April 2025

Fixed Effects, Lagged Dependent Variables, and Bracketing: Cautionary Remarks

Matei Demetrescu, Manuel Frondel, Lukas Tomberg, Colin Vance*

Suggested citation:
Demetrescu, M., Frondel, M., Tomberg, L., and Vance, C. (2025). "Fixed Effects, Lagged Dependent Variables, and Bracketing: Cautionary Remarks". *TRR 391 Working Paper #3*. DOI: 10.17877/DE290R-25433.

Version 1.0, April 2025

*Matei Demetrescu is a professor of Econometrics and Statistics at TU Dortmund University. Manuel Frondel, Lukas Tomberg, and Colin Vance are researchers at RWI – Leibniz Institute for Economic Research. In addition, Manuel Frondel is a professor of Energy Economics and Applied Econometrics at Ruhr University Bochum and Colin Vance is a professor of Quantitative Methods at Constructor University Bremen.
Corresponding author: lukas.tomberg@rwi-essen.de

Abstract

We investigate a bracketing property that purports to yield upper- and lower bounds on the treatment effects obtained from a fixed effects- and lagged dependent variable model. Referencing both analytical results and a Monte Carlo simulation, we explore the conditions under which the bracketing property holds, confirming this to be the case when the data generating process (DGP) is characterized by either unobserved heterogeneity or feedback effects from a lagged dependent variable (LDV). However, when the DGP is characterized by both features simultaneously, we find that bracketing of the treatment effect only holds under certain conditions – but not in general. Practitioners can nevertheless obtain the lower bound estimate by referencing a model that includes both fixed effects and a LDV. While the Nickell bias in the coefficient of the LDV is known to be of order $1/T$, we show that the Nickell-type bias in the estimator of the treatment effect is of order $1/T^2$.

Key words: Monte Carlo simulation, treatment effect, bounds, Nickell bias

1. Introduction

"So what's an applied guy to do?" Angrist and Pischke (2009, p. 245) pose this question in their discussion of the choice between the fixed-effects- and the lagged-dependent-variable model, addressing a dilemma that often confronts analysts of panel data. The dilemma comes down to this: The fixed-effects (FE) estimator, while controlling for unobserved, time-invariant heterogeneity that may be correlated with the regressors, relies on strict exogeneity.[1] One violation of this assumption is given by so-called feedback effects, wherein the past realization of the dependent variable affects the contemporaneous value of an explanatory or treatment variable. At the same time, while accommodating such feedback effects, the lagged-dependent-variable (LDV) estimator assumes that the unobserved heterogeneity is uncorrelated with the explanatory variables. For many causal questions, the assumptions underpinning either of the procedures are unlikely to be plausible (Angrist and Pischke, 2009, p. 245).

Drawing on the discussion in Guryan (2001), Angrist and Pischke (2009) suggest a way out of the dilemma by exploiting a bracketing property that is based on both the FE and the LDV estimates and allows practitioners to gauge the robustness of any estimate. The bracketing property is summarized by Guryan (2001, p. 55-56) as follows: If treatment is selected positively either on fixed characteristics or on a lagged dependent variable, then the FE estimator will yield a lower bound estimate while the LDV estimator will yield an upper bound estimate. Conversely, if treatment is selected negatively either on fixed characteristics or on a lagged dependent variable, then the FE estimator will yield an upper bound estimate while the LDV estimator will yield a lower bound estimate. The bracketing property thus suggests that by estimating both FE and LDV models, practitioners can bracket the causal effect of interest. Ding and Li (2019) have shown that this bracketing property of FE and LDV models also extends to general nonparametric settings.

The bracketing property has been increasingly applied in empirical papers in political science. For example, it is used by Keele et al. (2021), Tomberg et al. (2021), and Marsh (2023) to analyze voter behavior and by Keele et al. (2013) to analyze public spending. Yet it has also been used in other disciplines, for example in economics to analyze labor markets (Beckmann and Kräkel, 2022; Falk et al., 2018; Kampkötter and Sliwka, 2018), in biology to analyze the effect of biodiversity on the functioning of ecosystems (Dee et al., 2023), and in psychology by von Hippel (2022) to analyze the effect of replications on citations.[2] In addition, Angrist and Pischke's (2009) as well as Ding and Li's (2019) discussion of the properties of FE and LDV models is also referred to in the recent methodological syntheses of Roth et al. (2023) and Arkhangelsky and Imbens (2024).

The purpose of this article is to take a closer look at the conditions under which the bracketing property holds, both analytically and using a Monte Carlo simulation. Specifically, we demonstrate that the treatment effect cannot be bracketed when the unobserved hetero-

[1] Strict exogeneity is summarized by Wooldridge (2010) as $E(x'_{is}\varepsilon_{it}|\alpha_i) = 0, \quad s,t = 1,...,T$, i.e., the explanatory variable (x) is uncorrelated with all past and future disturbances (ε), conditional on fixed effects (α).

[2] We note that many of these studies do not rely exclusively on bracketing, but employ it as one of several identification strategies.

geneity is correlated with the regressors *and* the data generation process is characterized by feedback effects. We therefore conclude that the bracketing property may afford an unfounded picture of the true bounds and should be used with caution, particularly when the assumptions underpinning both models are questionable. In this instance, we advocate estimating an additional model that includes both fixed effects and a lagged dependent variable. Although this model does not allow for bracketing, the analytical results as well as the Monte Carlo evidence suggest that it can serve to identify the lower bound estimate of the treatment effect. Moreover, while the Nickell bias in the coefficient of the LDV is known to be of order $1/T$, we derive an expression for the "secondary" Nickell bias in the estimator of the treatment effect that is shown to be of order $1/T^2$ in our setup.

2. Background and Research Question

Our analysis picks up on the suggestion of Ding and Li (2019), who conclude with the open question of how the bracketing property would extend to a model that incorporates both fixed effects and a lagged dependent variable (we refer to this as the "FE-LDV model").[3] Although the FE-LDV model simultaneously controls for time-invariant unobserved heterogeneity and the feedback effect, its estimation requires stronger assumptions than either the FE or LDV models individually, a point recognized by Ding and Li (2019).

Indeed, the appeal of the bracketing property lies in its inherent promise to be an alternative to estimating the FE-LDV model, for which estimation will suffer from "Nickell bias" if the number of time periods is fixed (Nickell, 1981).[4] Under these circumstances, a Generalized Method of Moments approach in the style of the Arellano and Bond (1991) estimator (AB estimator) is a common way to obtain a consistent estimate of the treatment effect, τ.[5] However, as the AB estimator instruments the lagged dependent variable, $y_{i,t-1}$, with deeper lags of y_{it}, the method requires the availability of sufficient time periods. Additionally, the deeper lags of y_{it} need to be strong instruments for $y_{i,t-1}$, which may render the AB estimator inapplicable in many cases. Leszczensky and Wolbring (2022) discuss several of the challenges in applying the AB estimator, including downward bias in the face of a large number of moment conditions (Hsiao, 2022), weak instruments problems (Bun and Windmeijer, 2010), and poor finite-sample performance (Moral-Benito et al., 2019).

The bracketing property afforded by separate estimation of FE and LDV models avoids these complexities, and covers the circumstance when unobserved heterogeneity *or* feedback

[3] While Ding and Li (2019) employ a binary treatment and a two-period difference-in-differences estimator, we use a continuous treatment variable x_{it} and, similar to Angrist and Pischke (2009), a fixed-effects estimator with multiple time periods. While this difference does not matter for the bracketing property in general, we discuss in Section 7 a difference in the testability of the underlying assumptions between Ding and Li's (2019) and our setting.

[4] Nickell (1981) bias describes the bias that arises when a lagged dependent variable is included in a fixed effects model. In this case, the demeaned error term will be correlated with the lagged dependent variable, which leads to a bias.

[5] The cross-lagged panel model with FE (Moral-Benito, 2013; Moral-Benito et al., 2019) is another approach for obtaining consistent estimates. Leszczensky and Wolbring (2022) provide a comprehensive overview of alternative methods for addressing causal questions with panel data.

effects of the lagged outcome on the treatment are deemed threats to identification. This cuts to the issue raised by Ding and Li (2019): Their analysis, along with that of Guryan (2001) and Angrist and Pischke (2009), applies to the situation in which either the assumption underlying the FE model or the assumption underlying the LDV model is fulfilled. But given a concern for the validity of each assumption individually, the question arises as to the consequences for bracketing when the assumptions are simultaneously violated.

3. Biases in Case of Feedback and Endogeneity

An insightful case in which the assumptions of the FE model and the LDV model are violated is given by a DGP in which both the outcome variable y and the explanatory variable of interest x are functions of individual time-invariant factors α_i, i.e. fixed effects:

$$y_{it} = \delta_Y \alpha_i + \tau x_{it} + \varepsilon_{it}, \tag{1}$$

where x_{it} depends on the past realization $y_{i,t-1}$ of the outcome y, thereby constituting a violation of the strict exogeneity assumption due to a correlation between x_{it} and $\varepsilon_{i,t-1}$:

$$x_{it} = \delta_X \alpha_i + \rho y_{i,t-1} + u_{it}, \tag{2}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The effect of x on y, captured by τ, is the main parameter of interest. The coefficient ρ captures what we term the "feedback effect," and the condition $|\tau\rho| < 1$, together with suitable choice of the initial conditions y_{i0}, ensures stationarity, which we assume throughout our analysis:

Assumption 1 *Let y_{it} and x_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$, be generated as in* (1) *and* (2)*, where $|\tau\rho| < 1$ and the initial condition is given as*

$$y_{i0} = \frac{(\delta_Y + \tau\delta_X)\,\alpha_i}{1 - \tau\rho} + \frac{\tau u_{i0} + \varepsilon_{i0}}{\sqrt{1 - (\tau\rho)^2}}.$$

To focus on the essential aspects, the individual units are sampled independently but are distributed heterogeneously as implied by the following assumption.[6]

Assumption 2 *The disturbances $\varepsilon_{it} \sim (0, \sigma^2_{\varepsilon i})$ and $u_{it} \sim (0, \sigma^2_{ui})$, as well as the unobserved effects $\alpha_i \sim (\mu_\alpha, \sigma^2_\alpha)$, are mutually independent sequences of heterogeneous independent*

[6]The assumption is fairly general considering the type of data we model. Like the usual i.i.d assumption, it implies independent units – while allowing for different, unit-specific as well as time-specific specific distribution shapes. Moreover, it allows for different error variances across the panel. In exchange, we require a moment condition somewhat less general than finite variances: the imposed uniform (in i) moment boundedness allows for the application of suitable laws of large numbers (see the Technical Appendix C.1), and can best be interpreted in terms of an overall controlled propensity of the error distributions to generate only few large outliers.

random variables with uniformly bounded moments of order $2+\delta$ *for some* $\delta > 0$*, where* $\frac{1}{N}\sum_{i=1}^{N}\sigma_{ui}^2 \to \bar{\sigma}_u^2$ *and* $\frac{1}{N}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2 \to \bar{\sigma}_\varepsilon^2$ *as* $N \to \infty$*.*

It seems plausible that the simultaneous presence of fixed effects and a feedback effect as described in Equations (1) and (2) would emerge frequently. One example of this is the analysis of electoral outcomes: Much research is concerned with the question of how the regional variation of a factor x_{it} affects regional electoral outcomes (y_{it}), where the DGP may be subject to both feedback effects and unobserved heterogeneity. A feedback effect occurs when x_{it} – say, for example, regional unemployment or demographic composition – depends on past electoral outcomes ($y_{i,t-1}$) and thus on the existing political majorities in the time period between $t-1$ and t. Unobserved heterogeneity may be manifested in the form of unobservable time-constant factors, such as cultural or geographical features. In fact, three of the research papers mentioned in the introduction, in which FE and LDV models are estimated and reference is made to the bracketing property, are directly concerned with the analysis of electoral outcomes: Marsh (2023) analyzes the effects of traumatic events such as arson, mass shootings or natural disasters on voter turnout, Keele et al. (2021) analyze the impact of voting restrictions on voter registration, and Tomberg et al. (2021) analyze the impact of the presence of refugees on election outcomes.

If fixed effects and a feedback effect are simultaneously present, the estimate of τ from fitting either a FE model ($\ddot{y}_{it} = \tau\,\ddot{x}_{it} + error$ with $\ddot{\cdot}$ indicating within-transformation of the respective variable, e.g., $\ddot{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$) or a LDV model ($y_{it} = intercept + \omega y_{i,t-1} + \tau x_{it} + error$) will be biased. As derived for our setup in Appendix A.2, these biases (denoted by B_τ^{FE} and B_τ^{LDV}) are analytically tractable and can be simplified as summarized in the following

Proposition 1 *Under Assumptions 1 and 2, as* $N \to \infty$*, we have*

$$B_\tau^{FE} = -\rho\frac{1+\rho\tau}{T}\frac{\bar{\sigma}_\varepsilon^2}{\bar{\sigma}_u^2 + \rho^2\bar{\sigma}_\varepsilon^2} + O\left(\frac{1}{T}\right), \tag{3}$$

and

$$B_\tau^{LDV} = \frac{\delta_X\delta_Y\sigma_\alpha^2(\tau^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2)}{(\bar{\sigma}_u^2 + \delta_X^2\sigma_\alpha^2)(\rho^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2) + \bar{\sigma}_u^2\sigma_\alpha^2\frac{1+\tau\rho}{1-\tau\rho}(\delta_Y + \tau\delta_X)^2}. \tag{4}$$

where $O(\cdot)$ *denotes the order of magnitude.*[7]

Proof: *See Sections A.2.3 and A.2.4 in the Appendix.*

Let us for instance take the approximation in Equation (3) at face value. For the bracketing property to hold in this case, it is necessary that $sign(B_\tau^{FE}) \neq sign(B_\tau^{LDV})$, i.e., the signs of B_τ^{FE} and B_τ^{LDV} must always be in opposite directions for all combinations of different values for $\rho, \tau, \bar{\sigma}_\varepsilon^2, \bar{\sigma}_u^2, \sigma_\alpha, \delta_X$, and δ_Y in the DGP. Yet, it is clearly visible that the sign of

[7] Therefore, $O(1/T)$ stands for a vanishing quantity as T increases, and the approximation using just the leading term increases in precision as T increases.

B_τ^{FE} depends on the sign of ρ, while the sign of B_τ^{LDV} depends on the sign of $\delta_X \times \delta_Y$. Thus, as illustrated in Table 1, the bracketing property only holds if ρ and $\delta_X \times \delta_Y$ have the same signs. As discussed in Section 7, whether this is the case is likely to be difficult to determine in most practical applications.

If the signs are the opposite, then the bracketing property does not hold, i.e., both the FE and the LDV estimates of τ lie above or below the true value of τ. This means that the bracketing property fails given a DGP in which y and x are positively selected on the fixed effect α_i while the feedback effect is negative, i.e., the treatment is negatively selected on past realizations of the outcome. Moreover, the bracketing property fails given a DGP in which y is positively selected on the fixed effect, while x is negatively selected on the fixed effect and the feedback effect is positive.

Table 1: Illustration of the conditions under which the bracketing property holds, i.e., $sign(B_\tau^{FE}) \neq sign(B_\tau^{LDV})$

ρ	$\delta_X \times \delta_Y$	$sign(B_\tau^{FE}) \neq sign(B_\tau^{LDV})$
+	+	Yes
−	−	Yes
−	+	No
+	−	No

The following section confirms the predictions of Table 1 using Monte Carlo simulations.

4. Simulation Evidence

To illustrate these theoretical results, we conduct a Monte Carlo Simulation that demonstrates the performance of the FE and LDV estimators given a DGP containing fixed effects as well as feedback effects. As a comparison, we also investigate the performance of an OLS estimator ($y_{it} = intercept + \tau x_{it} + error$) as well as a FE-LDV estimator that includes both fixed effects and a lagged dependent variable ($\ddot{y}_{it} = \omega \ddot{y}_{i,t-1} + \tau \ddot{x}_{it} + error$)[8], which is known to suffer from Nickell bias.

As a starting point, we parameterize the DGP described by Equations (1) and (2) as follows: The fixed effect α_i is generated as a random variable drawn from a normal distribution with mean = 0 and standard deviation = 1 $[N(0,1)]$ once for each individual and remains constant over time. Then, ε, the error term affecting y, and u, the error term affecting x, are both i.i.d. and drawn from a $N(0,1)$ distribution. The starting values for the dynamic process are i.i.d. draws from a $N(0,1)$ distribution. Following Chudik and Pesaran (2019), we discard the first 50 simulated periods to avoid an influence of the starting values on

[8] As with the FE estimator, $\ddot{}$ symbolizes a within-transformation here; unlike the FE estimator, the mean values required for the transformation are formed here starting from period t=2 instead of t=1, since in common empirical applications there is no observation for the lagged dependent variable in the first period.

the simulation results, such that y_{0i} may be seen as being drawn from the stationary distribution.

Our parameterization of the DGP is intentionally very simplified to make the simulation results as comprehensible as possible. Therefore, the simulation results can only be interpreted in terms of the presence and direction of the bias. The reader should not over-interpret the magnitude of the bias in the simulated estimators, as the magnitude depends strongly on the underlying parameterization for which there is an infinite number of different possible combinations.

4.1. Evidence for Bracketing

To illustrate the bracketing relationship, we simulate scenarios in which either the assumptions underlying the LDV model or those underlying the FE model are fulfilled, with the results of these simulations being reported in Tables 2 and 3. Scenarios A to D in Table 2 are based on DGPs that contain a feedback effect of lagged outcomes $y_{i,t-1}$ on contemporaneous x_{it}, but there are no fixed effects. In this case, the LDV and OLS estimators are unbiased, as they only require contemporaneous exogeneity of x, which holds here. In contrast, the estimate of the FE model is biased owing to the violation of strict exogeneity due to the feedback effect. The estimate of the FE-LDV model is also biased due to two channels: the standard Nickell bias of order $1/T$ that applies to the estimator of the autoregressive coefficient, and, as we show below, a secondary Nickell bias of order $1/T^2$ that applies to the coefficients of the remaining explanatory variables.

Table 2: Monte Carlo Simulation results if the LDV model is correct, that is, $\rho \neq 0$ and there are no fixed effects: $\delta_X = 0 = \delta_Y$

	Data generating process				Estimates of τ				
Scenario	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV	Bracketing
A	1.00	0.00	0.50	0.00	1.00	0.89	1.00	0.95	Yes
B	1.00	0.00	-0.50	0.00	1.00	1.03	1.00	0.99	Yes
C	-1.00	0.00	0.50	0.00	-1.00	-1.03	-1.00	-0.99	Yes
D	-1.00	0.00	-0.50	0.00	-1.00	-0.89	-1.00	-0.95	Yes

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods. The data generating process is defined by Equations (1) and (2). The variables α_i, ε_{it}, u_{it} and y_{i1} are all i.i.d. draws from a normal distribution with mean = 0 and standard deviation = 1. To mitigate the potential influence of the starting value y_{i1} on the simulation results, we simulate 50 additional time periods and discard the first simulated 50 periods prior to estimating τ.

Next, the data underlying Scenarios E to H is generated by a DGP that includes a fixed effect that simultaneously influences y and x. As expected, the OLS estimator is biased in all cases, as ignoring fixed effects leads to omitted variable bias (Table 3). The LDV model suffers from the same bias. Conversely, the FE model eliminates this bias and yields correct estimates of the treatment effect. The FE-LDV model again suffers from Nickell bias.

Table 3: Monte Carlo Simulation results if the fixed effects model is correct, that is, $\delta_X \neq 0 \neq \delta_Y$ and there is no feedback effect: $\rho = 0$

	Data generating process				Estimates of τ				
Scenario	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV	Bracketing
E	1.00	0.50	0.00	0.50	1.20	1.00	1.14	0.98	Yes
F	1.00	-0.50	0.00	0.50	0.80	1.00	0.80	0.98	Yes
G	-1.00	0.50	0.00	0.50	-0.80	-1.00	-0.80	-0.98	Yes
H	-1.00	-0.50	0.00	0.50	-1.20	-1.00	-1.14	-0.98	Yes

See notes to Table 2.

To exemplify the validity of the bracket relationship, we employ the definition given by Angrist and Pischke (2009, p. 246) and focus on the results of Scenario B presented in Table 2 and the results on Scenario F reported in Table 3. According to Angrist and Pischke (2009), for positive treatment effects, $\tau > 0$, the bracketing property reads as follows: If the LDV-model "is correct, but you mistakenly use fixed effects, estimates of a positive treatment effect will tend to be too big. On the other hand, if [the fixed effects model] is correct and you mistakenly estimate an equation with lagged outcomes, [...] estimates of a positive treatment effect will tend to be too small." This definition is summarized in the following Table 4:

Table 4: Illustration of the direction of biases that lead to the bracketing property as described by Angrist and Pischke (2009)

LDV Model	FE Model	Estimated Model	Bias if $\tau > 0$
correct	incorrect	FE model	+
incorrect	correct	LDV model	-

Given the treatment effect estimate $\hat{\tau}_{LDV} = 0.80$ resulting from mistakenly estimating a LDV model while the FE model is correct (Scenario F in Table 3), and the alternative estimate $\hat{\tau}_{FE} = 1.04$ resulting from mistakenly estimating a FE model while the LDV model is correct (Scenario B in Table 2), the bracketing property holds, as claimed by Angrist and Pischke (2009):

$$\tau_{LDV} = 0.80 < \tau = 1 < \hat{\tau}_{FE} = 1.04.$$

We focused in this exemplary illustration of the bracketing property on scenarios in which the feedback effect is negative (Scenario B) or treatment is negatively selected on the fixed effect (Scenario F), because Angrist and Pischke (2009) use an example in which the treatment is a government-sponsored training program that targets individuals with poor labor market outcomes in the past. If contemporaneous labor market outcomes are the outcome variable of interest, such a selection process is represented by a negative feedback effect ($\rho < 0$, Scenario B) or a negative selection of treatment on unobservable 'ability' ($\delta_X < 0$, Scenario F), which may subsume factors such as intelligence that are largely constant across reasonable observation windows.

A more general definition of the bracketing property is provided in Guryan (2001, p. 55), who explicitly specifies how the bracketing property depends on the signs of ρ and δ_X: "if treatment is positively (negatively) selected on lagged outcomes, [that is, if $\rho > 0$ ($\rho < 0$),] the difference in-differences [or the fixed effects] estimator produces negatively (positively) biased estimates of the treatment effect." Moreover, "if treatment is positively (negatively) selected on fixed characteristics, [that is, if $\delta_X > 0$ ($\delta_X < 0$),] the estimator that controls for lagged outcomes produces positively (negatively) biased estimates of the treatment effect."

Based on this summary in Guryan (2001), the bracketing property is entirely confirmed by our simulation results, as presented in the following Table 5:

Table 5: Illustration of how the bracketing property manifests itself in our simulation results

	Scenario (Parameters)	Lower bound	τ	Upper bound	Scenario (Parameters)
$\tau > 0$:	A ($\rho > 0, \delta_X = 0$)	$\hat{\tau}_{FE} = 0.89$	$\tau = 1$	$\hat{\tau}_{LDV} = 1.14$	E ($\rho = 0, \delta_X > 0$)
	F ($\rho = 0, \delta_X < 0$)	$\hat{\tau}_{LDV} = 0.80$	$\tau = 1$	$\hat{\tau}_{FE} = 1.03$	B ($\rho < 0, \delta_X = 0$)
$\tau < 0$:	C ($\rho > 0, \delta_X = 0$)	$\hat{\tau}_{FE}$ = -1.03	$\tau = -1$	$\hat{\tau}_{LDV}$ = -0.80	G ($\rho = 0, \delta_X > 0$)
	H ($\rho = 0, \delta_X < 0$)	$\hat{\tau}_{LDV}$ = -1.14	$\tau = -1$	$\hat{\tau}_{FE}$ = -0.89	D ($\rho < 0, \delta_X = 0$)

4.2. Evidence against Bracketing if the FE-LDV Model holds True

Having established that our DGP is in line with the bracketing property of LDV and FE models, we now allow for both a fixed effect that influences y and x and a feedback of lagged outcomes $y_{i,t-1}$ on contemporaneous x_{it}. As predicted by the analytical results in Section 3, the figures in Table 6 illustrate that all estimates are biased and that the bracketing property no longer holds generally, although in Scenarios A, D, E, and H, the true effect lies within the FE and the LDV estimates.

Table 6: Results when there is a feedback effect of $y_{i,t-1}$ on x_{it} and a fixed effect simultaneously influences y_{it} and x_{it}

	Data generating process				Estimates of τ				
Scenario	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV	Bracketing
A	1.00	0.50	0.50	0.50	1.19	0.89	1.09	0.95	Yes
B	1.00	-0.50	0.50	0.50	0.87	0.89	0.80	0.95	No
C	1.00	0.50	-0.50	0.50	1.05	1.03	1.18	0.99	No
D	1.00	-0.50	-0.50	0.50	0.87	1.03	0.80	0.99	Yes
E	-1.00	0.50	0.50	0.50	-0.87	-1.03	-0.80	-0.99	Yes
F	-1.00	-0.50	0.50	0.50	-1.05	-1.03	-1.18	-0.99	No
G	-1.00	0.50	-0.50	0.50	-0.87	-0.89	-0.80	-0.95	No
H	-1.00	-0.50	-0.50	0.50	-1.19	-0.89	-1.09	-0.95	Yes

See notes to Table 2.

However, in Scenarios B, C, F and G, the true effect is not bracketed by the estimates of the FE and the LDV models, as these estimates are both either above or below the true value of τ. When examining the parameters entered into the DGP for these models, it becomes apparent that these are precisely the situations for which we predict the bracketing property not to hold based on the analytical expressions of the biases (see Table 1).

5. Deriving a Lower Bound from the FE-LDV Model

Our results so far show that the bracketing property afforded by the FE and the LDV models does not generally hold in the presence of unobserved heterogeneity *and* feedback effects. Hence, when there is a concern that the DGP is characterized by both these features simultaneously, relying on bracketing to identify the upper and lower bounds of the estimated treatment effect is ill-advised. Nevertheless, the results from the simulations suggest that it is at least possible to identify a lower bound estimate: Regardless of the DGP, the FE-LDV model always yields an estimate that is in absolute terms lower than the true coefficient, thus providing a lower bound of the true causal effect of x on y.

To confirm this pattern, we explored several DGPs with different parametrizations and found no instances in which the FE-LDV estimate is either larger than the true causal effect in absolute terms or in which it has a different sign than the true causal effect. These robustness tests are documented in Appendix B.1 and include variations in the intensity of the feedback effects (Appendix Table A3), the introduction of a dependence between the magnitude of the fixed effect (α_i) and the feedback effect (Appendix Table A4) and variations in the noise levels, i.e., the standard deviations of ϵ_{it}, α_t, u_{it} (Appendix Table A5). Apart from noting that none of these changes to the DGP alter the key findings from our main simulation specification, the potential to draw generalizable conclusions from our robustness tests is limited, as the behavior of the estimates is highly dependent on the particular parametrization. Nonetheless, in the cases where a direct comparison is possible, the response of our simulation results to changes in DGP proves to be consistent with the theoretical predictions from Equations (3) and (4) as well as Equation (6) derived below. For example, stronger feedback effects seem to be associated with a stronger downward bias in the FE-LDV estimate, but only when $sign(\tau) = sign(\rho)$, otherwise it is the other way around (Table A3), and we find that a relative increase in the standard deviation of u_{it} tends to reduce the bias of all estimators (Table A5).

In the following, we analytically derive the bias of the FE-LDV estimator of τ (denoted by B_T^{FE-LDV}) for any T. Starting with $T = 3$, the bias is given by[9]

Proposition 2 *Under the Assumptions of Proposition 1 it holds as $N \to \infty$ that*

$$B_3^{FE-LDV} = -\frac{\frac{1}{4}\tau\bar{\sigma}_u^2\bar{\sigma}_\varepsilon^2}{\bar{\sigma}_u^2\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\tau^2\rho^2} - \frac{1}{4}\tau^2\bar{\sigma}_u^4} = -\tau\frac{1}{\tau^2\frac{\bar{\sigma}_u^2}{\bar{\sigma}_\varepsilon^2}\left(\frac{4}{1-\tau^2\rho^2}-1\right)+\frac{4}{1-\tau^2\rho^2}}.$$

[9]Note that $T = 3$ is the smallest possible number of time periods given our notation. This is because the first period ($t = 1$) is the initial period for which we have an observation for y but not for the lagged dependent variable, and one further needs at least two periods to apply the within transformation.

Proof: *See Section A.2.2 in the Appendix.*

This expression highlights that the bias of the FE-LDV estimator is negative for positive τ and positive for negative τ, respectively. In turn, this hinges on the condition that $|B_\tau^{FE\text{-}LDV}| \leq |\tau|$. Should this not be met, the FE-LDV estimator would provide an estimate of τ that has the wrong sign, which would critically limit the applicability of the FE-LDV estimator to provide a lower bound estimate. The condition $|B_\tau^{FE\text{-}LDV}| \leq |\tau|$ translates into:

$$\tau^2 \frac{\bar{\sigma}_u^2}{\bar{\sigma}_\varepsilon^2} \left(\frac{4}{1-\tau^2\rho^2} - 1 \right) + \frac{4}{1-\tau^2\rho^2} > 1, \tag{5}$$

which holds true generally, since $\tau^2 \frac{\bar{\sigma}_u^2}{\bar{\sigma}_\varepsilon^2} > 0$ and $\frac{4}{1-\tau^2\rho^2} > 4 \; \forall \, |\tau\rho| < 1$.

While these considerations may be repeated for any T, the corresponding expressions become less tractable as T increases, and we consider an approximation following the lines of Nickell (1981). The result takes the following form:

Proposition 3 *Under the Assumptions of Proposition 1 it holds as $N \to \infty$ that*

$$B_\tau^{FE\text{-}LDV} = -\frac{\tau}{(T-1)^2} \frac{\bar{\sigma}_\varepsilon^2}{\tau^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2} \frac{1+\tau\rho}{1-\tau\rho} + O\left(\frac{1}{T^2}\right). \tag{6}$$

Proof: *See Section A.2.2 in the Appendix.*

Interestingly, this bias in the estimator of the effect τ vanishes at rate $1/(T-1)^2$, which is an order of magnitude faster than the Nickell bias in the FE-LDV estimator of the autoregressive (AR) coefficient, which is itself inversely proportional to T. This somewhat surprising finding is, to the best of our knowledge, new, and it is of immediate relevance to practitioners interested in identifying the causal effect of a treatment. The standard expression for Nickell bias, which is of order $1/T$, applies only to the AR coefficient, and does not in general carry over to other explanatory variables, a point often neglected in the applied literature.[10] Equation (6) provides an approximate expression for a secondary Nickell bias that applies to the remaining explanatory variables. We provide a discussion of its relation to Nickell's original result in Appendix A.2.2.

Notwithstanding its speed of convergence, the bias of the FE-LDV estimator is negative for positive τ and positive for negative τ, respectively, and thus suggests that the FE-LDV estimator may indeed provide a lower bound estimate for $T > 3$ as well. The condition $|B_\tau^{FE\text{-}LDV}| \leq |\tau|$ simplifies to:

$$\tau\rho \leq 1 - \frac{2}{1 + (T-1)^2 \frac{\tau^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{\bar{\sigma}_\varepsilon^2}}, \tag{7}$$

[10] For example, Acemoglu et al. (2019) (p. 59) conclude that the bias emerging from a FE-LDV model of the effect of democratization on GDP growth is $1/T$.

which almost certainly holds as long is T is large and the time series of y is stationary, i.e., $-1 < \tau\rho < 1$.

What about if T is small? We have seen that the condition $|B_{\tau}^{FE\text{-}LDV}| \leq |\tau|$ holds for $T = 3$ irrespective of the actual persistence. Moreover, we simulate scenarios with low T and $\tau\rho$ close to unity in Table A1 in the Appendix. In these simulations the bias of the FE-LDV model does not even get close to the size of τ, supporting the analytical results.

To illustrate the behavior of the estimators at different T and in the presence of fixed effects and feedback effects, we present our simulation results using the parametrization from Table 6 over different values of T in Appendix B.5 (Figure A2). It can be seen that the biases of the OLS and LDV models are not affected by increasing T, so these estimators do not approach the true effect as T increases. In contrast, both the FE and FE-LDV estimators converge towards the true effect. While they are approximately equidistant from the true value of τ at $T = 3$, the bias of the FE-LDV estimator vanishes quickly and is relatively close to the true effect from $T = 20$ in our simulated case, while this convergence takes longer for the FE model. In line with our theoretical results, the FE-LDV model converges to the correct value from below in all simulated cases, while the FE model sometimes converges from below (Scenarios A and B) and sometimes from above (Scenarios C and D).

Concluding, when suspecting the simultaneous presence of fixed effects and a violation of strict exogeneity due to a feedback effect, referencing the FE-LDV model avails the practitioner with a conservative estimate of the treatment effect, including when T is small – although in this case a substantial underestimation of the treatment effect is to be expected. Indeed, we would caution against trying to discern an upper bound, even in cases where one seems apparent. In Scenario A from Table 6, for example, where the FE-LDV estimate is above the FE- and below the LDV estimates, the LDV model cannot be interpreted unambiguously as an upper bound. This is because all estimates could be downward biased while the FE-LDV estimate lies coincidentally between the FE and LDV estimates, a particular case that is presented in Scenario A in Table A2 of the Appendix. Similarly, if the FE-LDV estimate is below both the FE and the LDV estimate, as in Scenario C from Table 6, it again cannot be guaranteed that the LDV estimate is an upper bound estimate, as all estimates may be downward biased, a case presented in Scenario B from Table A2 in the Appendix.

6. The Implications of State Dependence and Nonstationarity

One natural extension of our DGP is to include state dependence in y_{it}, i.e., to allow y_{it} to be directly influenced by $y_{i,t-1}$. This changes Equation (1) to:

$$y_{it} = \delta_Y \alpha_i + \pi y_{i,t-1} + \tau x_{it} + \varepsilon_{it}. \tag{8}$$

This DGP requires the stationarity condition $|\pi + \tau\rho| < 1$, reflecting that y_{it} now depends on $y_{i,t-1}$ both directly (via $\pi y_{i,t-1}$ in (8)) and indirectly (via the feedback to τx_{it}).

The simulations based on this DGP (setting π to 0.2), are presented in Appendix Tables A3, A4, and A5. Note that the OLS and FE models are still estimated as static models, i.e., not including $y_{i,t-1}$ among the regressors. There are several important takeaways from this extension of the DGP: First, in the case where there are no fixed effects but a feedback effect of $y_{i,t-1}$ on x_{it}, the inclusion of a direct effect of $y_{i,t-1}$ on y_{it} in the DGP leads to bias in the OLS estimates (Table A3), which were unbiased before (Table 2). This is intuitive, as the omission of $y_{i,t-1}$, which is now correlated with y_{it} as well as x_{it}, leads to omitted variable bias due to the correlation between the error term and x_{it}.

Second, somewhat surprisingly, given state dependence, the static fixed effects estimator is also biased if there are fixed effects but no feedback effect of $y_{i,t-1}$ on x_{it} (compare Table A4 with Table 3). The intuition for this bias was recently laid out by Klosin (2024): Omitting the lagged dependent variable in the estimating equation leads to $\varepsilon_{i,t+1}$ being a function of y_{it} and thus of x_{it}. This violates the assumption of strict exogeneity required by the FE model. The upshot, demonstrated in Scenarios F and G in Table A4, shows that it is not even necessary for both fixed effects and a feedback effect to be present for the bracketing property to be violated. Even with a completely random x_{it}, the bracketing property of FE and LDV can fail if the DGP contains state dependence.

A further result that calls for caution, especially when interpreting the FE model, can be found in Appendix Figure A3. There we show the behavior of the different estimators in the face of feedback effects, fixed effects and state dependence over different observation periods T. Scenarios A - in which FE and LDV would actually bracket the true effect in the absence of state dependence - and B show that the FE estimator changes its sign with increasing T. This means that while the FE and LDV models bracket the true effect in Scenario A at low T, they fail to do so at higher T. Analogously, in Scenario B the two estimators do not bracket the true effect at low T, while they do so at higher T. This supports our conclusion that the presence of state dependence makes a bracketing approach profoundly unreliable.

As in the case without state dependence, the estimates of the FE-LDV model are consistently below the true effect in absolute values. We corroborate these with theoretical results established in the Technical Appendix and summarized below.

Proposition 4 *Let y_{it} be generated as in (8) with x_{it} given by (2), where*

$$y_{i0} = \frac{(\delta_Y + \tau\delta_X)\,\alpha_i}{1-(\pi+\tau\rho)} + \frac{\tau u_{i0} + \varepsilon_{i0}}{\sqrt{1-(\pi+\tau\rho)^2}}$$

s.t. $|\pi + \tau\rho| < 1$ and u_{it}, ε_{it} and α_i obey Assumption 2. Then, it holds as $N \to \infty$ that

$$B_T^{FE} = \rho\frac{\bar{\sigma}_\varepsilon^2\left(-\frac{1}{T}\frac{1}{1-(\pi+\tau\rho)}\right)}{\left(\rho^2\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-(\pi+\tau\rho)^2}+\bar{\sigma}_u^2\right)}+\pi\left(\rho\frac{\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-(\pi+\tau\rho)^2}\left(1-\frac{1}{T}\frac{1+(\pi+\tau\rho)}{1-(\pi+\tau\rho)}\right)}{\left(\rho^2\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-(\pi+\tau\rho)^2}+\bar{\sigma}_u^2\right)}+\tau\frac{\bar{\sigma}_u^2\left(-\frac{1}{T}\frac{1}{1-(\pi+\tau\rho)}\right)}{\left(\rho^2\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-(\pi+\tau\rho)^2}+\bar{\sigma}_u^2\right)}\right)+O\left(\frac{1}{T}\right),$$

$$B_T^{LDV} = \frac{\delta_X \delta_Y \sigma_\alpha^2 \left(\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2\right)}{\left(\bar{\sigma}_u^2 + \delta_X^2 \sigma_\alpha^2\right)\left(\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2\right) + \bar{\sigma}_u^2 \sigma_\alpha^2 \frac{1+(\pi+\tau\rho)}{1-(\pi+\tau\rho)} \left(\delta_Y + \tau\delta_X\right)^2}.$$

and

$$B_T^{FE-LDV} = -\frac{\tau}{(T-1)^2} \frac{\bar{\sigma}_\varepsilon^2}{\left(\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2\right)} \frac{1+(\pi+\tau\rho)}{1-(\pi+\tau\rho)} + O\left(\frac{1}{T^2}\right).$$

Proof: *See Section A.3 in the Appendix.*

Compared to the baseline case without explicit state dependence, the expressions for B_T^{LDV} and B_T^{FE-LDV} do not change significantly, in fact the only difference is that now $\pi + \tau\rho$ is the relevant persistence parameter. Only the expression of B_T^{FE} changes compared to Proposition 1, given that the FE model is misspecified when there is direct state dependence and the bias now has two sources, namely the omission of $y_{i,t-1}$ as well as the lack of strict exogeneity of x_{it}. Notwithstanding this change, the conclusions about the validity of the bounding behavior do not change.

A further relevant extension consists in relaxing the stationarity assumption. In particular, we consider nonstationarity in form of a unit root, $\pi + \tau\rho = 1$. Detailed derivations pertaining to this case are provided in the Technical Appendix, and the following proposition shows that the bias of the FE-LDV estimator behaves qualitatively the same even under nonstationarity.

Proposition 5 *Let y_{it} be generated as in (8) with x_{it} given by (2), $\pi + \tau\rho = 1$, and u_{it}, ε_{it} and α_i obey Assumption 2. Then, it holds as $N \to \infty$ that*

$$B_T^{FE-LDV} = -\tau \frac{\bar{\sigma}_\varepsilon^2}{\frac{1}{3}\left(\left(\delta_Y + \tau\delta_X\right)^2 \sigma_\alpha^2 (T-1)T + \left(\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2\right)(4T-3)\right) - \tau^2 \bar{\sigma}_u^2}$$

Proof: *See Section A.4 in the Appendix.*

Like above, the direction of the bias is opposite to that of the effect τ. The exact expression of the bias does differ from the stationary case. One interesting difference is that the unobserved heterogeneity α_i does play a role in the limit. In fact, the rate at which the bias vanishes as $T \to \infty$ depends on whether unobserved heterogeneity is present in the model. If this is not the case (i.e., $\delta_Y = 0$ and $\delta_X = 0$), the rate of convergence is $1/T$ as opposed to the rate of convergence of $1/T^2$ that we found for all other settings. On the other hand, the initial conditions y_{i0} need not be restricted in any way, as they are eliminated by the fixed effects transformation. Nevertheless, it is evident that the bias is smaller in magnitude than the effect for any $T \geq 3$ and the overall conclusion does not change.

Appendix B.3 contains simulation results for the case of a unit root ($\rho = 1,\ \tau = 1,\ \pi = 0$) and confirms the theoretical results: The FE-LDV model continues to provide an estimate that is consistently lower in absolute terms than the true effect. In addition, when comparing, for example, Scenario A in Table A9 to Scenario A in Table A10, i.e., once without and once with an effect of unobserved heterogeneity on X and Y at constant T,

the simulation results confirm that the bias of the FE-LDV model is actually larger if there is no effect of unobserved heterogeneity.

7. Practical Implications

Returning to the initial question "so what's an applied guy to do?" (Angrist and Pischke, 2009, p. 245), the key implication of our results is that the bracketing property is not a panacea in situations where empirical researchers are concerned with confounding by unobserved time-constant heterogeneity as well as violations of the assumption of strict exogeneity, for example because of feedback effects or state dependence. This is because the bracketing property is only reliable if either strict exogeneity holds *or* there is no confounding by unobserved time-constant heterogeneity. Given the likely uncertainty surrounding this issue, it may be possible to avail diagnostics tests that allow the researcher to rule out one source of bias or the other. For example, unobserved time-constant heterogeneity can be explored using tests in the spirit of Hausman's (1978) specification test (Frondel and Vance, 2010), while statistical tests for the strict exogeneity assumption are proposed by Wooldridge (2010, p. 285). The outcome of such an exercise may clearly point to the preferability of the FE or LDV model, as would be the case if *either* the strict exogeneity assumption is violated *or* a fixed effects model is required. Bracketing in this case is unnecessary.

Another case in which there is clarity about the path forward is the two-period setting with a binary treatment covered by Ding and Li (2019). The authors propose a test in which the cumulative distribution function (CDF) of the pre-treatment outcome variable for the treatment and the control group are plotted against each other (see Keele et al., 2021 for an application). If one CDF is monotonically above/below the other (stochastic monotonicity), Ding and Li's (2019) results imply that the bracketing property is likely hold.

However, if one moves away from the two-period difference-in-differences setting with binary treatment, this approach reaches its limits. As our analysis has shown, with multiple periods, bracketing only works if the selection of the independent variable of interest x depends in the same direction, i.e., positively or negatively, on lagged outcomes and the relevant omitted time constant variables ($sign(\rho) = sign(\delta_X \times \delta_Y)$, see Table 1), and if there is no state dependence. One example in which these conditions might be met is the study of Beckmann and Kräkel (2022), who, among other identification strategies, estimate FE and LDV models to bracket the effects of work autonomy (x) on work engagement (y). In this case, we might tell a story about an unobservable fixed factor such as *ability* that makes the sign of $\delta_X \times \delta_Y$ positive, coupled with the expectation that past engagement positively affects today's autonomy, making ρ positive as well. Under this circumstance, the potential selection of x on fixed effects should go in the same direction as the potential selection of x on past outcomes, supporting bracketing.

Whether such a story is convincing is, of course, open to interpretation. In our view,

an analyst will typically be hard-pressed to make an airtight case for bracketing.[11] We therefore recommend that authors transparently discuss their theoretical considerations on the possible selection process, taking into account our insights from Section 3, and indicate in which of the cases listed in Table 1 they see their setting. If these theoretical considerations lead to an ambiguous result, as we believe will often be the case, authors should not rely on the validity of a bracketing approach. Instead, authors should pay particular attention to the FE-LDV model, which provides a conservative reference point, yielding a lower bound of the true effect, even if the nature of the selection process is unclear. Moreover, we show that the bias of the FE-LDV estimator of the explanatory variables decreases at rate $1/T^2$ (see Proposition 3), so that the results of the FE-LDV model converge relatively fast towards the true effect as T increases.

An empirical example in which the presence of fixed effects, state dependence, and feedback effects would hypothetically render the result of a bracketing approach misleading can be found in Acemoglu et al. (2019). The authors estimate the effect of democratization on economic growth using a panel of 175 countries from 1960 to 2010. The authors implement several specifications but do not reference bracketing. A key observation, demonstrated descriptively in their Figure 1, is that democratization is often preceded by a decline in GDP, which suggests the presence of a feedback effect with $\rho < 0$. Moreover, they argue persuasively that "democracies differ from nondemocracies in unobserved characteristics, such as institutional, historical, and cultural aspects" (p. 49). Together, these considerations lead the authors to estimate dynamic panel data models, including a FE-LDV model, throughout the paper. Using the replication files of Acemoglu et al. (2019), we estimate the results if the authors had tried to bracket the true effect with FE and LDV models, focusing on the simplest specification in column (1) of their Table 2. The estimate of the effect of democratization on log GDP per capita would be 0.457 (0.300) in the LDV model and -10.112 (4.316) in the FE model (standard errors in parentheses).[12] In contrast, their reported and replicable result of the FE-LDV model is 0.973 (0.294), while the result of their preferred specification in column (3), which includes four lags of the dependent variable, is 0.787 (0.226), both falling outside the range of the bracket.

Throughout most of this article, we have assumed contemporaneous exogeneity, i.e., zero correlation between x_{it} and ε_{it} for all t and conditional on accounting for fixed effects. If this condition is not met, conventional approaches to causal inference (e.g., instrumental variables), possibly also in combination with a bracketing approach, should be considered. Furthermore, the dynamic structure that we have adopted in our DGPs is relatively simple, but should cover a wide range of applications. Nevertheless, there may be cases in which deeper lags of the dependent or explanatory variables may be part of the DGP (e.g., Acemoglu et al., 2019). A generalization of our results to these cases and a discussion of the implications for the interpretability of the estimated coefficients, e.g., with respect to short-run and long-run effects (Beck and Katz, 2011; Keele and Kelly, 2006), is beyond the scope of this article.

[11] The relationship between traumatic events and voter turnout analyzed by Marsh (2023), or the effect of biodiversity on productivity analyzed by Dee et al. (2023), are examples in which various confounders can influence the selection mechanism in very complex ways.

[12] See Table A11 in the Appendix for details.

8. Conclusion

This article has explored the conditions under which a fixed effects- and a lagged dependent variable model can be used to bracket a causal effect of interest. We draw two conclusions. First, bracketing works in the presence of *either* unobserved heterogeneity *or* violations of strict exogeneity, but not both simultaneously. Second, even when it is unclear whether the data generation process is determined by fixed effects and/or a lagged dependent variable, our results indicate that the analyst can at least identify a conservative lower bound estimate of the treatment effect with a model that includes both features. Of particular relevance to the case of short panels, we provide an approximate expression for a secondary Nickell bias of this treatment effect and the remaining explanatory variables, which is of order $1/T^2$.

We recommend that before employing a bracketing approach, practitioners should first use diagnostic tests to investigate whether selection based on time-constant unobservable variables and violations of strict exogeneity are present simultaneously. If only one or the other is present, a researcher should be able to obtain an unbiased estimate from either an FE or LDV model. If there is reason to expect simultaneous selection based on time-constant unobservable variables and violations of strict exogeneity, for example due to a feedback effect, researchers should form a theoretical expectation about the direction of these two selection effects. If they are in the same direction, e.g., positive selection based on a time-constant confounder and positive selection based on past outcomes, a bracketing approach may be valid. However, as there exists no test for this, such a consideration is always associated with uncertainty, and sometimes the conceivable selection effects can be so complex that no meaningful theoretical expectation is possible. Furthermore, the presence of state dependence, i.e., a direct effect of past outcomes on current outcomes, can jeopardize the validity of a bracketing approach by introducing an additional source of bias in the FE model.

These considerations lead us to regard the bracketing approach with FE and LDV models to be a risky strategy in most cases. When used, we recommend complementing the approach with an FE-LDV model. This model provides an estimate that converges reliably from below (i.e. from 0) to the true effect in the scenarios considered at a rate of $1/T^2$. One exception is the case in which there is no selection on time-constant unobservables and at the same time there is a unit root. In this case, the estimator continues to converge from below towards the true value, but with the convergence rate $1/T$. Furthermore, we recommend that the estimation of an FE-LDV model should completely replace a bracketing approach when the number of observed periods is sufficiently large. Our simulations show that the number of periods should be at least 20 for the FE-LDV model to provide a good approximation of the true effect.

Acknowledgments

This research was funded in the course of TRR 391 *Spatio-temporal Statistics for the Transition of Energy and Transport* (520388526) by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and within Kopernikus Project Ariadne 2 by the Federal Ministry of Education and Research (BMBF). The authors would like to thank Marco Horvath and Rubén Ruiz Rufino for helpful comments.

Appendix

A. Formal derivation of biases

A.1. Estimators and tools

In the **FE-LDV model**, one accounts for both unobserved heterogeneity and feedback and we fit

$$\ddot{y}_{it} = \hat{\omega}^{FE-LDV}\ddot{y}_{i,t-1} + \hat{\tau}^{FE-LDV}\ddot{x}_{it} + error,$$

where " $\ddot{}$ " signifies that the variables have been within-transformed. To maintain consistency, we take the within transformation to be applied onto the r.h.s. *after* lagging y_{it}, that is, we compute for $t = 2, \ldots, T$ the individually demeaned variables $\ddot{x}_{it} = x_{it} - \frac{1}{T-1}\sum_{t=2}^{T} x_{it} = x_{it} - \bar{x}_i$, $\ddot{y}_{it} = y_{it} - \frac{1}{T-1}\sum_{t=2}^{T} y_{it} = y_{it} - \bar{y}_i$ and $\ddot{y}_{i,t-1} = y_{i,t-1} - \frac{1}{T-1}\sum_{t=2}^{T} y_{i,t-1} = y_{i,t-1} - \bar{y}_{i-}$.[13] This ensures e.g. that

$$\ddot{y}_{it} = \tau\ddot{x}_{it} + \ddot{\varepsilon}_{it}, \qquad t = 2, \ldots, T,$$

where $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \frac{1}{T-1}\sum_{t=2}^{T}\varepsilon_{it}$, and, with $\ddot{u}_{it} = u_{it} - \frac{1}{T-1}\sum_{t=2}^{T} u_{it}$,

$$\ddot{x}_{it} = \rho\ddot{y}_{i,t-1} + \ddot{u}_{it}.$$

The estimated effect is given by

$$\hat{\tau}^{FE-LDV} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{x}_{it}\ddot{y}_{it} - \sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{x}_{it}\ddot{y}_{i,t-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{y}_{i,t-1}^2\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{y}_{i,t-1}\ddot{y}_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{x}_{it}^2 - \sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{x}_{it}\ddot{y}_{i,t-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{y}_{i,t-1}^2\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{y}_{i,t-1}\ddot{x}_{it}}.$$

It will be convenient to stack all observations unit-wise, resulting in

$$\hat{\tau}^{FE-LDV} = \frac{\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_i - \frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_i}{\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i - \frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i}$$

where $\ddot{\boldsymbol{y}}_i$ stacks $\ddot{y}_{i2}, \ldots, \ddot{y}_{iT}$, $\ddot{\boldsymbol{x}}_i$ stacks $\ddot{x}_{i2}, \ldots, \ddot{x}_{iT}$ and $\ddot{\boldsymbol{y}}_{i-}$ stacks $\ddot{y}_{it-1}$ for $t = 2, \ldots, T$.

In the **FE model**, the regression does not control for a lagged dependent variable and one uses the standard within-transformed data, $\ddot{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\ddot{x}_{it} = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}$ for $t = 1, \ldots, T$,

$$\ddot{y}_{it} = \hat{\tau}^{FE}\ddot{x}_{it} + error,$$

[13] We allow for a slight inconsistency in notation: the lagged $\ddot{y}_{it}$ does not equal $\ddot{y}_{i,t-1}$ because the two variables are demeaned in a different manner, $\bar{y}_i = \frac{1}{T-1}\sum_{t=2}^{T} y_{it} \neq \frac{1}{T-1}\sum_{t=2}^{T} y_{i,t-1} = \bar{y}_{i-}$. This does not have any impact on the derivations here, though, since only $\ddot{y}_{i,t-1}$ appears as a lag variable in the derivations.

leading to the somewhat simpler expression

$$\hat{\tau}^{FE} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}\,\ddot{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}^{2}} = \frac{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\,\ddot{\boldsymbol{y}}_i}{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\,\ddot{\boldsymbol{x}}_i}.$$

For the **LDV model** one fits

$$y_{it} = intercept + \hat{\omega}^{LDV} y_{it-1} + \hat{\tau}^{LDV} x_{it} + error$$

without individual dummy variables. Therefore, we have

$$\hat{\tau}^{LDV} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{x}_{it}\tilde{y}_{it} - \sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{x}_{it}\tilde{y}_{i,t-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{y}_{i,t-1}^{2}\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{y}_{i,t-1}\tilde{y}_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{x}_{it}^{2} - \sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{x}_{it}\tilde{y}_{i,t-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{y}_{i,t-1}^{2}\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{y}_{i,t-1}\tilde{x}_{it}},$$

where " $\tilde{\cdot}$ " signifies that the variables have been demeaned using the "full-panel" means, $\tilde{x}_{it} = x_{it} - \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T} x_{it} = x_{it} - \bar{x}$, $\tilde{y}_{it} = y_{it} - \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it} = y_{it} - \bar{y}$, $\tilde{y}_{i,t-1} = y_{it-1} - \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1} = y_{i,t-1} - \bar{y}_{-}$ etc.[14] Then, with $\tilde{\boldsymbol{y}}_i$ stacking $\tilde{y}_{it}$, $\tilde{\boldsymbol{x}}_i$ stacking $\tilde{x}_{it}$ and $\tilde{\boldsymbol{y}}_{i-}$ stacking $\tilde{y}_{it-1}$ for $t = 2, \ldots, T$, we have analogously that

$$\hat{\tau}^{LDV} = \frac{\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{y}}_i - \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{y}}_{i-}\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}'\tilde{\boldsymbol{y}}_{i-}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}'\tilde{\boldsymbol{y}}_i}{\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{x}}_i - \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{y}}_{i-}\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}'\tilde{\boldsymbol{y}}_{i-}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}'\tilde{\boldsymbol{x}}_i}.$$

All estimators are given in terms of sample averages of independent variables, and we resort to a suitable law of large numbers to analyze the respective limits of the estimators and the implied asymptotic biases. To this end, we use the following result.

Theorem 1 *If $\boldsymbol{z}_i$ are uniformly integrable, independent random vectors, then $\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{z}_i - \mathrm{E}(\boldsymbol{z}_i)) \xrightarrow{p} \mathbf{0}$;[15] furthermore, if $\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(\boldsymbol{z}_i)$ converges to a fixed $\overline{\mathrm{E}(\boldsymbol{z}_i)}$, then $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_i \xrightarrow{p} \overline{\mathrm{E}(\boldsymbol{z}_i)}$.*

We therefore only need to work out the corresponding expectations and argue that their average converges to a fixed number, which is done in the Technical Appendix (Appendix C).

[14] There is again a slight inconsistency in notation, since lagging $\tilde{y}_{it}$ does not give $\tilde{y}_{it-1}$ because the two variables undergo a different demeaning scheme, $\bar{y} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it} \neq \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1} = \bar{y}_{-}$.

[15] This is proved in Chow (1971). To use it in our setup, note first that uniform boundedness of moments of any order $\eta > 1$ implies uniform integrability, and second that quantities like $\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}$ do indeed have uniformly bounded such moments – since the Cauchy-Schwarz inequality implies that any expectation of the form $\mathrm{E}\left(\left|\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right|^{\eta}\right)$ is uniformly bounded for some $\eta > 1$ if $\ddot{\boldsymbol{x}}_i$ and $\ddot{\boldsymbol{y}}_{i-}$ have uniformly bounded moments of some order larger than 2, and is indeed the case here given that $\ddot{\boldsymbol{x}}_i$ and $\ddot{\boldsymbol{y}}_{i-}$ are essentially linear combinations of errors fulfilling this moment condition.

A.2. Derivations for the baseline setup

A.2.1. Data generating process

The data generating process is

$$y_{it} = \delta_Y \alpha_i + \tau x_{it} + \varepsilon_{it}, \qquad t = 1, \ldots, T, \tag{A1}$$

exhibiting feedback and fixed effects correlated with the regressor,

$$x_{it} = \delta_X \alpha_i + \rho y_{it-1} + u_{it}, \qquad t = 1, \ldots, T. \tag{A2}$$

Following the assumptions in the main text, the error vectors $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ and $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iT})'$, $i = 1, \ldots, N$, are mutually independent zero-mean sequences with unit-specific variances and free of correlation in the time dimension, i.e.

$$\mathrm{Cov}\left(\boldsymbol{\varepsilon}_i\right) = \sigma^2_{\varepsilon i} \mathbf{I}_T \qquad \mathrm{Cov}\left(\boldsymbol{u}_i\right) = \sigma^2_{ui} \mathbf{I}_T,$$

where $\mathbf{I}_T$ stands for the identity matrix of order T. The fixed effects α_i are independent across the panel, independent of the errors $\boldsymbol{\varepsilon}_j$ and $\boldsymbol{u}_j$ for all $1 \leq j \leq N$, and satisfy $\mathrm{E}\left[\alpha_i\right] = \mu_\alpha$ and $\mathrm{Var}\left[\alpha_i\right] = \sigma^2_\alpha$. We furthermore impose stationarity requirements (which imply $|\tau\rho| < 1$ in this formulation of the DGP) on the initial conditions y_{i0}; see the main text and Technical Appendix C.2.1).

A.2.2. Asymptotic bias of the FE-LDV estimator

Since within-transforming Equation (A1) ensures that

$$\ddot{\boldsymbol{y}}_i = \tau \ddot{\boldsymbol{x}}_i + \ddot{\boldsymbol{\varepsilon}}_i,$$

it follows in the FE-LDV model that

$$\hat{\tau}^{FE-LDV} = \tau + \frac{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i - \frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-} \left(\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right)^{-1} \frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i}{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i - \frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-} \left(\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right)^{-1} \frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i}. \tag{A3}$$

We refer to the Technical Appendix (Appendix C.2.2) for the derivation of the limits of these sample cross-product moments.

Let, for ease of notation, the limits of the average expectations be denoted by an overline, e.g. $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right] := \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}'\ddot{\boldsymbol{\varepsilon}}\right]}$.

To first understand the *direction* of the asymptotic bias of $\hat{\tau}$, it suffices to examine the sign of the limit of the numerator of the term capturing the estimation error in Equation (A3), i.e.,

$$\ddot{M}_T := \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]} - \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]} \cdot \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}}.$$

This expression simplifies in our setup, since within-transforming Equation (A2) implies

$$\ddot{\boldsymbol{x}}_i = \rho \ddot{\boldsymbol{y}}_{i-} + \ddot{\boldsymbol{u}}_i,$$

such that

$$\ddot{M}_T = \rho \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{\varepsilon}}_i\right]} - \frac{\left(\rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\right]}\right)\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]}}.$$

Note now that $\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{\varepsilon}}_i\right]}$ must be zero since $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{u}_i$ are zero-mean mutually independent vectors, and we have

$$\begin{aligned}
\ddot{M}_T &= \rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]} - \frac{\rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]}} - \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]}} \\
&= -\frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]}}.
\end{aligned}$$

The denominator of the expression of the asymptotic bias,

$$\ddot{N}_T = \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{x}}_i\right]} - \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{x}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]}},$$

is positive by construction, just like $\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right]$; we furthermore have from the Technical Appendix (Appendix C.2.2) that

$$\begin{aligned}
\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\right]} &= (T-1)\,\tau\bar{\sigma}_u^2\left(-\frac{1}{T-1}\frac{1}{1-\tau\rho}\right) + O(1) \\
\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]} &= (T-1)\,\bar{\sigma}_\varepsilon^2\left(-\frac{1}{T-1}\frac{1}{1-\tau\rho}\right) + O(1)
\end{aligned}$$

with $O(\cdot)$ denoting the order of magnitude (and therefore $O(1)$ standing for a bounded quantity as T increases). We may therefore state (approximately) that

$$\operatorname{sign}\left(\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\right]}\right) = -\operatorname{sign}(\tau)$$

and

$$\operatorname{sign}\left(\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]}\right) = -1.$$

Summing up, the sign of the leading term of the asymptotic bias of the estimated effect is the opposite of the sign of τ. This implies that the bias is in the opposite direction of the actual effect, just as supported by the simulation results.

To justify the statement from the main text on the FE-LDV estimate as a lower bound for the effect, we need to derive the *magnitude* of the asymptotic bias. Taking into account

that

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]} = \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]} = \rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]},$$

the asymptotic bias, say B_T^{FE-LDV}, is then given by

$$B_T^{FE-LDV} = \frac{\ddot{M}_T}{\ddot{N}_T} = -\frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} - \left(\rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]}\right)^2},$$

where

$$\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right] = \rho^2\,\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right] + 2\rho\,\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right] + \mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i\right]$$

and therefore

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} = \rho^2\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} + 2\rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i\right]},$$

such that

$$B_T^{FE-LDV} = -\frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} - \left(\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]}\right)^2}.$$

Using again the results from the Technical Appendix (Appendix C.2.2), this results e.g. for $T = 3$ in

$$B_3^{FE-LDV} = -\frac{\frac{1}{4}\tau\bar{\sigma}_u^2\bar{\sigma}_\varepsilon^2}{\bar{\sigma}_u^2\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\tau^2\rho^2} - \frac{1}{4}\tau^2\bar{\sigma}_u^4} = -\tau\frac{1}{\tau^2\frac{\bar{\sigma}_u^2}{\bar{\sigma}_\varepsilon^2}\left(\frac{4}{1-\tau^2\rho^2}-1\right)+\frac{4}{1-\tau^2\rho^2}}.$$

The case $\tau = 0$ leads (almost trivially) to a zero bias, while, for $\tau \neq 0$, the condition $\left|B_3^{FE-LDV}\right| < |\tau|$ translates into

$$\tau^2\frac{\bar{\sigma}_u^2}{\bar{\sigma}_\varepsilon^2}\left(\frac{4}{1-\tau^2\rho^2}-1\right)+\frac{4}{1-\tau^2\rho^2} > 1,$$

which is easily seen to hold true since $\tau^2\frac{\bar{\sigma}_u^2}{\bar{\sigma}_\varepsilon^2} > 0$ and $\frac{4}{1-\tau^2\rho^2} > 4\ \forall\,|\tau\rho| < 1$.

Since the expression of B_T^{FE-LDV} gets involved already for $T = 4$, we resort to the approximations derived in the Technical Appendix (Appendix C.2.2), which are quite precise for larger T. We have in this case

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]} = \tau\bar{\sigma}_u^2\bar{\sigma}_\varepsilon^2\left(\frac{1}{1-\tau\rho}\right)^2 + O(1);$$

furthermore,

$$\begin{aligned}\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} &= (T-1)^2\,\bar{\sigma}_u^2\left(1-\frac{1}{T-1}\right)\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-(\tau\rho)^2}\left(1-\frac{1}{T-1}\frac{1+\tau\rho}{1-\tau\rho}\right)+O(T)\\ &= (T-1)^2\,\frac{\bar{\sigma}_u^2\left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)}{1-(\tau\rho)^2}+O(T),\end{aligned}$$

and, finally,

$$\left(\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{u}}_i\right]}\right)^2 = \tau^2\bar{\sigma}_u^4\left(\frac{1}{1-\rho\tau}\right)^2 + O\left(1\right).$$

Summing up,

$$B_T^{FE-LDV} = -\frac{\tau}{\left(T-1\right)^2}\frac{\bar{\sigma}_\varepsilon^2}{\left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)}\frac{1+\tau\rho}{1-\tau\rho} + O\left(\frac{1}{T^2}\right).$$

We notice that the bias decreases in T^2 (i.e. faster than the Nickell bias in the autoregressive coefficient estimator which is inversely proportional to T) and is thus negligible for large T. It is even negligible for moderate T as long as the absolute value of $\tau\rho$ is not too close to unity. When focusing on the leading term as an approximation for the asymptotic bias, one may learn more about the behaviour of the bias; note however that the approximation likely deteriorates when $|\tau\rho|$ is very close to unity, and is entirely invalid when $\tau\rho = \pm 1$.

In order to justify the magnitude of the estimated effect as a lower bound for the actual effect (while maintaining the direction of the effect), it should hold that

$$\left|B_T^{FE-LDV}\right| \leq |\tau|,$$

or, with $\frac{1+\tau\rho}{1-\tau\rho}$ being positive under the stationarity restriction $-1 < \rho\tau < 1$,

$$\frac{\bar{\sigma}_\varepsilon^2}{\left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)}\frac{1+\tau\rho}{1-\tau\rho} < \left(T-1\right)^2.$$

This is equivalent to

$$\tau\rho < 1 - \frac{2}{1+\left(T-1\right)^2\left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)/\bar{\sigma}_\varepsilon^2}$$

Since $\left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)/\bar{\sigma}_\varepsilon^2 > 1$, the above inequality is satisfied if $\tau\rho < 1 - \frac{2}{1+(T-1)^2}$, irrespective of the variances of the error components. Several such upper bounds on persistence are plotted in Figure A1 as a function of the time dimension and of the signal-to-noise ratio $\tau^2\bar{\sigma}_u^2/\bar{\sigma}_\varepsilon^2$.

We note that the effect itself plays an important role, since a non-zero τ (i.e. a positive signal-to-noise ratio) shifts this bound closer to unity (the simulations in Table A1 suggest that the lower-bound interpretation does not experience difficulties even when $|\tau\rho| = 0.99$ and $T = 3$, where $1 - \frac{2}{1+(T-1)^2} = 0.6$).

Finally, it would be of interest to align our results with those of Nickell (1981) when considering additional r.h.s. variables. His discussion assumes exogenous regressors. Denote by $\hat{\omega}$ the fitted autoregressive coefficient, and note that, in the model without state dependence, $\omega = 0$. Then, in our notation, his Eq. (24) implies that

$$\operatorname{plim}\hat{\tau}^{FE-LDV} - \tau = -\operatorname{plim}\frac{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{y}}_{i-}}{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{x}}_i}\operatorname{plim}\left(\hat{\omega}-\omega\right) + \operatorname{plim}\frac{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{\varepsilon}}_i}{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{x}}_i},$$

where, in Nickell's 1981 setup with exogeneity of $\boldsymbol{x}_i$, $\operatorname{plim}\left(\hat{\omega}-\omega\right) = O(1/(T-1))$ and the

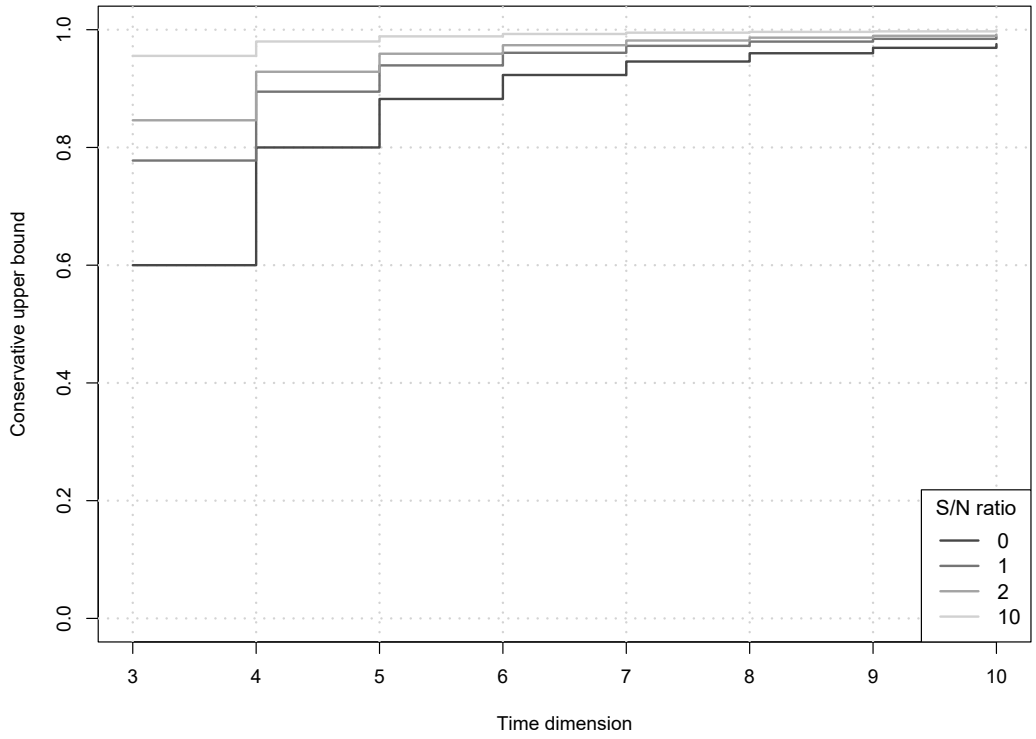

Figure A1: The conservative upper bound for the persistence $\tau\rho$

2nd term on the r.h.s. is 0. This leads to his Eq. (26) which seems to suggest that there are cases where B_τ^{FE-LDV} is of order $O(1/T)$.

To explain this seeming discrepancy, we distinguish in our setup between the cases without and with feedback.

Take first the case without feed-back in our setup, i.e. set $\rho = 0$. Then Nickell's 1981 Eq. (26) does follow, since $\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i \xrightarrow{p} 0$. However,

$$\operatorname{plim} \frac{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}}{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i} = \operatorname{plim} \frac{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{y}}_{i-}}{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i} \xrightarrow{p} \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i\right]}} = O\left(\frac{1}{T-1}\right)$$

such that the resulting magnitude of the bias of $\hat{\tau}^{FE-LDV}$ is, as derived above, indeed $1/(T-1)^2$.

Take now the case with a feed-back effect, $\rho \neq 0$. In this case,

$$\operatorname{plim} \frac{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i}{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i} \xrightarrow{p} \frac{\rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\rho^2\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} + 2\rho\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right] + \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i\right]}} \neq 0,$$

in fact the r.h.s. can be seen to be $O(1/(T-1))$. Since in this case

$$\operatorname{plim} \frac{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}}{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i} = \frac{\rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]}}{\rho^2\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} + 2\rho\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]} + \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{u}}_i\right]}} = O(1),$$

and still $\operatorname{plim}(\hat{\omega}-\omega) = O(1/(T-1))$, Eq. (26) of Nickell (1981) is not valid under our setup. Rather, it can be seen after some algebra that we recover the result of Propositions 2 and 3 from the main text as follows. Write

$$\operatorname{plim}\hat{\tau}^{FE-LDV} - \tau = \frac{1}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]}}\left(-\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\operatorname{plim}(\hat{\omega}-\omega) + \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}\right)$$

where

$$\begin{aligned}
\hat{\omega}^{FE-LDV} &= \frac{\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i - \frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i\ddot{\boldsymbol{\varepsilon}}_i}{\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i - \frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}}. \\
&\xrightarrow{p} \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}}.
\end{aligned}$$

Then,

$$\begin{aligned}
B_\tau^{FE-LDV} &= -\frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]}\left(\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\right)} \\
&\quad + \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]}\left(\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\right)} \\
&= \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]}\left(\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\right)} \\
&= \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}}.
\end{aligned}$$

Exploiting like above $\ddot{\boldsymbol{x}}_i = \rho\ddot{\boldsymbol{y}}_{i-} + \ddot{\boldsymbol{u}}_i$, we obtain immediately that

$$\begin{aligned}
B_\tau^{FE-LDV} &= \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}} \\
&= \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} - \overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{x}}_i\right]}\cdot\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}}
\end{aligned}$$

where the denominator is easily checked to be $O((T-1)^2)$. At the same time, the numerator is $O(1)$ since both $\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{y}}_{i-}\right]}$ and $\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\right]}$ are $O(1)$ and $\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{\varepsilon}}_i\right]} = 0$ due to the mutual independence of the two error sequences u_{it} and ε_{it}. This confirms the result on B_τ^{FE-LDV}

derived above.

A.2.3. Asymptotic bias of the FE estimator

In what concerns the FE regression, we obtain analogously to the FE-LDV case that

$$\hat{\tau}^{FE} = \tau + \frac{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{\varepsilon}}_i}{\frac{1}{N}\sum_{i=1}^{N} \ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{x}}_i} \xrightarrow{p} \tau + \rho \frac{\overline{\mathrm{E}\left[\, \ddot{\boldsymbol{y}}_{i-}' \ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\, \ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{x}}_i\right]}}, \tag{A4}$$

where we used the fact that $\ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{\varepsilon}}_i = \rho\, \ddot{\boldsymbol{y}}_{i-}' \ddot{\boldsymbol{\varepsilon}}_i + \ddot{\boldsymbol{u}}_i' \ddot{\boldsymbol{\varepsilon}}_i$ with $\mathrm{E}\left[\, \ddot{\boldsymbol{u}}_i' \ddot{\boldsymbol{\varepsilon}}_i\right] = \mathbf{0}$ since $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{u}_i$ are zero-mean mutually independent.

Therefore, the direction of the bias depends directly on the sign of ρ, in fact it is the negative sign of ρ since $\operatorname{sign}\left(\overline{\mathrm{E}\left[\, \ddot{\boldsymbol{y}}_{i-}' \ddot{\boldsymbol{\varepsilon}}_i\right]}\right) = -1$. The magnitude follows from the results in the Technical Appendix (Appendix C.2.2),

$$\begin{aligned} B_T^{FE} = \rho \frac{\overline{\mathrm{E}\left[\, \ddot{\boldsymbol{y}}_{i-}' \ddot{\boldsymbol{\varepsilon}}_i\right]}}{\overline{\mathrm{E}\left[\, \ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{x}}_i\right]}} &= -\rho \frac{\bar{\sigma}_\varepsilon^2 \left(\frac{1}{1-\tau\rho}\right) + O(1)}{T\left(\rho^2 \frac{\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{1-(\rho\tau)^2} + \bar{\sigma}_u^2\right) + O(1)} \\ &= -\rho \frac{1+\rho\tau}{T} \frac{\bar{\sigma}_\varepsilon^2}{\bar{\sigma}_u^2 + \rho^2 \bar{\sigma}_\varepsilon^2} + O\left(\frac{1}{T}\right). \end{aligned}$$

The FE bias vanishes as T increases of course, albeit slower than the FE-LDV bias, which is inversely proportional to T^2. The same caveats arise, in that the quality of the approximation deteriorates for small T and $|\tau\rho|$ close to unity. In any case, note that the magnitude of the bias is not "under control" since ρ could be large if τ is small in magnitude under our assumptions, the stationarity restriction being $|\tau\rho| < 1$; in fact ρ could be any real number if $\tau = 0$ as there would be no feed-back in this extreme case.

A.2.4. Asymptotic bias of the LDV estimator

Note first that, not surprisingly, omitting individual-unit dummies does not remove the fixed effects, in fact we obtain from Equations (A1) and (A2) that

$$\begin{aligned} \tilde{\boldsymbol{y}}_i &= \boldsymbol{\delta}_Y \tilde{\alpha}_i + \tau \tilde{\boldsymbol{x}}_i + \tilde{\boldsymbol{\varepsilon}}_i \\ \tilde{\boldsymbol{x}}_i &= \boldsymbol{\delta}_X \tilde{\alpha}_i + \rho \tilde{\boldsymbol{y}}_{i-} + \tilde{\boldsymbol{u}}_i, \end{aligned}$$

where $\tilde{\alpha}_i = \alpha_i - \bar{\alpha}$ with $\bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i$, and $\boldsymbol{\delta}_X$ and $\boldsymbol{\delta}_X$ are $(T-1)$-vectors having δ_X and δ_Y as entries (i.e. $\boldsymbol{\delta}_X = \boldsymbol{\iota}\delta_X$ with $\boldsymbol{\iota}$ a vector of ones). This and the Frisch-Waugh-Lovell theorem imply in turn that

$$\hat{\tau} = \tau + \frac{\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i' \left(\tilde{\boldsymbol{\varepsilon}}_i + \boldsymbol{\delta}_Y \tilde{\alpha}_i\right) - \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{y}}_{i-} \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}' \tilde{\boldsymbol{y}}_{i-}\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}' \left(\tilde{\boldsymbol{\varepsilon}}_i + \boldsymbol{\delta}_Y \tilde{\alpha}_i\right)}{\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{x}}_i - \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i' \tilde{\boldsymbol{y}}_{i-} \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}' \tilde{\boldsymbol{y}}_{i-}\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i-}' \tilde{\boldsymbol{x}}_i}. \tag{A5}$$

We note that $\hat{\tau}$ is invariant to μ_α since all quantities are demeaned, so we may assume, without loss of generality, that $\mu_\alpha = 0$. This implies $\mathrm{E}[y_{it}] = \mathrm{E}[x_{it}] = 0$.

We now argue that demeaning does not have an asymptotic impact on the probability limit of $\hat{\tau}$. Take e.g.

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{y}}_{i-} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{x}\boldsymbol{\iota})'\left(\boldsymbol{y}_{i-} - \bar{y}_-\boldsymbol{\iota}\right) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{y}_{i-} - (T-1)\,\bar{x}\bar{y}_-$$

where $\bar{x}$ and $\bar{y}_-$ are both $O_p\left(N^{-1/2}\right)$, as a straightforward calculation of the variances shows. Therefore

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}_i'\tilde{\boldsymbol{y}}_{i-} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{y}_{i-} + O_p\left(\frac{T}{N}\right).$$

To put things in perspective, note that $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{y}_{i-} = O_p(T)$. Moreover, $\bar{\alpha} = O_p\left(N^{-1/2}\right)$ as well, and it is easily shown that $\bar{\varepsilon} = O_p\left(T^{-1/2}N^{-1/2}\right)$ and $\bar{u} = O_p\left(T^{-1/2}N^{-1/2}\right)$, such that the analogous result may be derived for all involved cross-product moments in Equation (A5). Summing up, for fixed T,

$$\hat{\tau} = \tau + \frac{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'(\boldsymbol{\varepsilon}_i + \boldsymbol{\delta}_Y\alpha_i) - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{y}_{i-}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y}_{i-}'(\boldsymbol{\varepsilon}_i + \boldsymbol{\delta}_Y\alpha_i)}{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{y}_{i-}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y}_{i-}'\boldsymbol{x}_i} + O_p\left(\frac{1}{N}\right)$$

such that the asymptotic bias B_T^{LDV} is given by the ratio of

$$\tilde{M}_T = \overline{\mathrm{E}\left[\boldsymbol{x}_i'(\boldsymbol{\varepsilon}_i + \boldsymbol{\delta}_Y\alpha_i)\right]} - \frac{\overline{\mathrm{E}\left[\boldsymbol{x}_i'\boldsymbol{y}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'(\boldsymbol{\varepsilon}_i + \boldsymbol{\delta}_Y\alpha_i)\right]}}{\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right]}}$$

and

$$\tilde{N}_T = \overline{\mathrm{E}\left[\boldsymbol{x}_i'\boldsymbol{x}_i\right]} - \frac{\overline{\mathrm{E}\left[\boldsymbol{x}_i'\boldsymbol{y}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'\boldsymbol{x}_i\right]}}{\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right]}}.$$

Plugging in $\boldsymbol{x}_i = \boldsymbol{\delta}_X\alpha_i + \rho\boldsymbol{y}_{i-} + \boldsymbol{u}_i$, we obtain thanks to the linearity of the expectation operator,

$$B_T^{LDV} = \frac{\tilde{M}_T}{\tilde{N}_T} = \frac{\overline{\mathrm{E}\left[(\boldsymbol{u}_i + \boldsymbol{\delta}_X\alpha_i)'(\boldsymbol{\varepsilon}_i + \boldsymbol{\delta}_Y\alpha_i)\right]}\cdot\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right]} - \overline{\mathrm{E}\left[(\boldsymbol{u}_i + \boldsymbol{\delta}_X\alpha_i)'\boldsymbol{y}_{i-}\right]}\cdot\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'(\boldsymbol{\varepsilon}_i + \boldsymbol{\delta}_Y\alpha_i)\right]}}{\overline{\mathrm{E}\left[(\boldsymbol{u}_i + \boldsymbol{\delta}_X\alpha_i)'(\boldsymbol{u}_i + \boldsymbol{\delta}_X\alpha_i)\right]}\cdot\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right]} - \left(\overline{\mathrm{E}\left[\boldsymbol{y}_{i-}'(\boldsymbol{u}_i + \boldsymbol{\delta}_X\alpha_i)\right]}\right)^2}.$$

Using the results from the Technical Appendix (Appendix C.2.3), we have that

$$\begin{aligned} B_T^{LDV} &= \frac{\delta_X\delta_Y\sigma_\alpha^2\left(\frac{(\delta_Y+\tau\delta_X)^2}{(1-\omega)^2}\sigma_\alpha^2 + \frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\omega^2}\right) - \delta_X\delta_Y\sigma_\alpha^2\frac{(\delta_Y+\tau\delta_X)^2}{(1-\omega)^2}\sigma_\alpha^2}{(\bar{\sigma}_u^2 + \delta_X^2\sigma_\alpha^2)\left(\frac{(\delta_Y+\tau\delta_X)^2}{(1-\omega)^2}\sigma_\alpha^2 + \frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\omega^2}\right) - \delta_X^2\sigma_\alpha^2\frac{(\delta_Y+\tau\delta_X)^2}{(1-\omega)^2}\sigma_\alpha^2} \\ &= \frac{\delta_X\delta_Y\sigma_\alpha^2(\tau^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2)}{(\bar{\sigma}_u^2 + \delta_X^2\sigma_\alpha^2)(\tau^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2) + \bar{\sigma}_u^2\sigma_\alpha^2\frac{1+\tau\rho}{1-\tau\rho}(\delta_Y + \tau\delta_X)^2}, \end{aligned}$$

which confirms the direction of the bias (and also the dependence on all involved quanti-

ties).

A.3. State-dependent model

A.3.1. Data generating process

The data generating process is now

$$y_{it} = \delta_Y \alpha_i + \pi y_{i,t-1} + \tau x_{it} + \varepsilon_{it}, \qquad t = 1, \ldots, T, \tag{A6}$$

with the regressors behaving as in Eq. (A2) above. The estimators remain the same.

As we argue in the Technical Appendix (Appendix C.3), the data generating mechanism of y_{it} from the baseline case is recovered by setting $\omega = \pi + \tau\rho$. We may therefore build on the corresponding limits for the cross-product sample moments.

A.3.2. Asymptotic bias of the FE-LDV estimator

The expressions obtained in Section C.2.2 are valid and we obtain the expression of the bias for the state-dependent case by replacing $\tau\rho$ by $\pi + \tau\rho$,

$$B_T^{FE-LDV} = -\frac{\tau}{(T-1)^2} \frac{\bar{\sigma}_\varepsilon^2}{(\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2)} \frac{1 + (\pi + \tau\rho)}{1 - (\pi + \tau\rho)} + O\left(\frac{1}{T^2}\right).$$

The discussion on the sign and the magnitude of this bias remain valid (with $\pi + \tau\rho$ replacing $\tau\rho$ of course) such that the FE-LDV estimator still offers a lower bound for the magnitude of the effect.

A.3.3. Asymptotic bias of the FE estimator

Here the expression of the bias does change in an essential manner since the FE model is misspecified whenever $\pi \neq 0$. We have for

$$\hat{\tau}^{FE} - \tau = \frac{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}_i' \left(\ddot{\boldsymbol{\varepsilon}}_i + \pi \ddot{\boldsymbol{y}}_{i-}' \right)}{\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{x}}_i} \xrightarrow{p} \rho \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}' \ddot{\boldsymbol{\varepsilon}}_i \right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{x}}_i \right]}} + \pi\rho \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}' \ddot{\boldsymbol{y}}_{i-} \right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{x}}_i \right]}} + \pi \frac{\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i' \ddot{\boldsymbol{y}}_{i-} \right]}}{\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i' \ddot{\boldsymbol{x}}_i \right]}},$$

such that

$$B_T^{FE} = \rho \frac{\bar{\sigma}_\varepsilon^2 \left(-\frac{1}{T} \frac{1}{1-\omega} \right)}{\left(\rho^2 \frac{\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{1-\omega^2} + \bar{\sigma}_u^2 \right)} + \pi \left(\rho \frac{\frac{\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{1-\omega^2} \left(1 - \frac{1}{T} \frac{1+\omega}{1-\omega} \right)}{\left(\rho^2 \frac{\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{1-\omega^2} + \bar{\sigma}_u^2 \right)} + \tau \frac{\bar{\sigma}_u^2 \left(-\frac{1}{T} \frac{1}{1-\omega} \right)}{\left(\rho^2 \frac{\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{1-\omega^2} + \bar{\sigma}_u^2 \right)} \right) + O\left(\frac{1}{T}\right).$$

We notice two bias components: The first is due to the feed-back effect to the regressor, whereas the second is due to the presence of the lagged dependent variable in the model.

The latter can be decomposed in two components as well, one that yet again depends on the feed-back effect, while the other depends on the actual effect of the regressor.

A.3.4. Asymptotic bias of the LDV estimator

As we show in the Technical Appendix (Appendix C.3), the data generating mechanism of y_{it} from the baseline case is recovered by setting $\omega = \pi + \tau\rho$. The arguments regarding the negligibility of the demeaning step hold just like in the baseline setup, and we ultimately obtain

$$B_T^{LDV} = \frac{\delta_X \delta_Y \sigma_\alpha^2 \left(\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2\right)}{\left(\bar{\sigma}_u^2 + \delta_X^2 \sigma_\alpha^2\right)\left(\tau^2 \bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2\right) + \bar{\sigma}_u^2 \sigma_\alpha^2 \frac{1+(\pi+\tau\rho)}{1-(\pi+\tau\rho)} \left(\delta_Y + \tau\delta_X\right)^2}.$$

A.4. Nonstationarity

Let us now consider the case of a unit root ($\pi + \tau\rho = 1$) as a specific form of deviation from stationarity.

Like before, the FE transform implies for $i = 1, \ldots, N$ that

$$\ddot{\boldsymbol{y}}_i = \pi \ddot{\boldsymbol{y}}_{i-} + \tau \ddot{\boldsymbol{x}}_i + \ddot{\boldsymbol{\varepsilon}}_i \qquad \text{and} \qquad \ddot{\boldsymbol{x}}_i = \rho \ddot{\boldsymbol{y}}_{i-} + \ddot{\boldsymbol{u}}_i,$$

where, for $t = 2, \ldots, T$, $\boldsymbol{u}_i$ and $\boldsymbol{\varepsilon}_i$ stack u_{it} and ε_{it}, and $\boldsymbol{y}_i$ and $\boldsymbol{y}_{i-}$ stack y_{it} and y_{it-1}, and $\ddot{\cdot}$ stands for individually demeaned quantities, say $\ddot{y}_{it} = y_{it} - \frac{1}{T-1}\sum_{t=2}^{T} y_{it}$ and $\ddot{y}_{it-1} = y_{it-1} - \frac{1}{T-1}\sum_{t=2}^{T} y_{it-1}$.

With initial conditions y_{i0}, we have

$$x_{i1} = \delta_X \alpha_i + \rho y_{i0} + u_{i1}$$

and, with $\tau\rho + \pi = 1$, it follows that

$$y_{i1} = \delta_Y \alpha_i + \pi y_{i0} + \tau x_{i1} + \varepsilon_{i1} = y_{i0} + \left(\delta_Y + \tau\delta_X\right)\alpha_i + \left(\tau u_{i1} + \varepsilon_{i1}\right).$$

Iterating, we obtain for all $1 \leq t \leq T$

$$y_{it} = y_{i0} + t\left(\delta_Y + \tau\delta_X\right)\alpha_i + \sum_{s=1}^{t}\left(\tau u_{is} + \varepsilon_{is}\right).$$

Let now $\boldsymbol{\iota}_{1:t}$ denote the T-vector whose first t elements are 1 and the rest are 0. We thus have for all $1 \leq t \leq T$

$$y_{it} = y_{i0} + \left(\delta_Y + \tau\delta_X\right)\boldsymbol{\iota}_{1:t}'\boldsymbol{\iota}_T\alpha_i + \tau\boldsymbol{\iota}_{1:t}'\boldsymbol{u}_{T,i} + \boldsymbol{\iota}_{1:t}'\boldsymbol{\varepsilon}_{T,i},$$

where $\boldsymbol{\iota}_T = \boldsymbol{\iota}_{1:T}$ stands for a T-vector of ones and $\boldsymbol{u}_{T,i}$ and $\boldsymbol{\varepsilon}_{T,i}$ stack u_{it} and ε_{it} for $t = 1, \ldots, T$; these are distinct from $\boldsymbol{u}_i$ and $\boldsymbol{\varepsilon}_i$ which stack u_{it} and ε_{it} for $t = 2, \ldots, T$ only.

Furthermore, denote by $\boldsymbol{\Theta}_-$ the $(T-1)\times T$ matrix stacking $\boldsymbol{\iota}'_{1:t}$ for $t=1,\dots,T-1$, i.e.

$$\boldsymbol{\Theta}_- = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ 1 & 1 & 1 & \ddots & & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & & \ddots & 0 & 0 \\ 1 & 1 & 1 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

As such, we obtain, with $\boldsymbol{\iota}_{T-1}$ a $T-1$-vector of ones,

$$\boldsymbol{y}_{i-} = y_{i0}\boldsymbol{\iota}_{T-1} + (\delta_Y + \tau\delta_X)\,\alpha_i\boldsymbol{\Theta}_-\boldsymbol{\iota}_T + \tau\boldsymbol{\Theta}_-\boldsymbol{u}_{T,i} + \boldsymbol{\Theta}_-\boldsymbol{\varepsilon}_{T,i}.$$

and, with $\mathbf{Q}_{T-1} = \mathbf{I}_{T-1} - \frac{1}{T-1}\boldsymbol{\iota}_{T-1}\boldsymbol{\iota}'_{T-1}$ the usual "demeaning matrix" (which, recall, is symmetric and idempotent by construction),

$$\ddot{\boldsymbol{y}}_{i-} = \mathbf{Q}_{T-1}\boldsymbol{y}_{i-} = (\delta_Y + \tau\delta_X)\,\alpha_i\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{\iota}_T + \tau\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{u}_{T,i} + \mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{\varepsilon}_{T,i}.$$

We note in passing that, unlike in the stationary case, the fixed-effects transformation does not remove the cumulated unobserved effect $t\,(\delta_Y + \tau\delta_X)\,\alpha_i$ (which effectively acts like a trend in this model) from either y_{it} or y_{it-1}, and therefore the unobserved effects α_i will still play a role in the limit. The initial conditions y_{i0} are however washed out by the fixed-effects transform.

Like in Eqn. A3, we have that the asymptotic bias of $\hat{\tau}^{FE-LDV}$ is given by the probability limit of a certain ratio,

$$B_T^{FE-LDV} = \operatorname*{plim}_{N\to\infty} \frac{\sum \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{\varepsilon}}_i - \sum \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{y}}_{i-}\left(\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right)^{-1}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i}{\sum \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{x}}_i - \sum \ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{y}}_{i-}\left(\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right)^{-1}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{x}}_i},$$

which simplifies due to $\ddot{\boldsymbol{x}}_i = \rho\ddot{\boldsymbol{y}}_{i-} + \ddot{\boldsymbol{u}}_i$ to

$$B_T^{FE-LDV} = \operatorname*{plim}_{N\to\infty} \frac{\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{\varepsilon}}_i\frac{1}{N}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-} - \frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\cdot\frac{1}{N}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i}{\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\frac{1}{N}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-} - \left(\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\right)^2}.$$

Since, as $N\to\infty$, $\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{\varepsilon}}_i \xrightarrow{p} 0$ and $\frac{1}{N}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}$ is bounded, we obtain

$$B_T^{FE-LDV} = -\frac{\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\cdot\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i}{\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum \ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-} - \left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{y}}_{i-}\right)^2}.$$

Analyzing the sample averages, we know from Section C.2.2 that

$$\frac{1}{N}\sum \ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i \xrightarrow{p} \overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\right]} = (T-1)\,\bar{\sigma}_u^2\left(1-\frac{1}{T-1}\right) = (T-2)\,\bar{\sigma}_u^2;$$

the other terms require however a nonstationarity-specific analysis which we provide in

the following.

Note first that

$$\begin{aligned}\frac{1}{N}\sum \ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{y}}_{i-} &= \frac{1}{N}\sum \ddot{\boldsymbol{u}}_i'\left(\left(\delta_Y+\tau\delta_X\right)\alpha_i\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{\iota}_T+\tau\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{u}_{T,i}+\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{\varepsilon}_{T,i}\right)\\ &\xrightarrow{p} \tau\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{u}_{T,i}\right]};\end{aligned}$$

this is because α_i and $\boldsymbol{u}_i$, as well as $\boldsymbol{u}_i$ and $\boldsymbol{\varepsilon}_i$, are mutually independent. With $\mathbf{S}$ the $(T-1)\times T$ selector matrix, selecting $\boldsymbol{u}_i=(u_{i2},\ldots,u_{iT})$ from $\boldsymbol{u}_{T,i}$,

$$\mathbf{S}=\begin{pmatrix}0&1&0&0&\cdots&\cdots&0\\0&0&1&0&\cdots&\cdots&0\\0&0&0&1&\ddots&&0\\\vdots&\vdots&\vdots&\vdots&\ddots&\ddots&\vdots\\0&0&0&0&&\ddots&0\\0&0&0&0&\cdots&\cdots&1\end{pmatrix}$$

s.t. $\boldsymbol{u}_i=\mathbf{S}\boldsymbol{u}_{T,i}$, we have

$$\begin{aligned}\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}_i'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{u}_{T,i}\right]} &= \overline{\mathrm{E}\left[\boldsymbol{u}_{T,i}'\mathbf{S}'\mathbf{Q}_{T-1}'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{u}_{T,i}\right]}=\overline{\mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{u}_{T,i}\boldsymbol{u}_{T,i}'\mathbf{S}'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\right)\right]}\\ &= \bar{\sigma}_u^2\,\mathrm{tr}\left(\mathbf{S}'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\right).\end{aligned}$$

Similarly, with $\boldsymbol{\varepsilon}_i=\mathbf{S}\boldsymbol{\varepsilon}_{T,i}$, it follows that

$$\frac{1}{N}\sum\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{\varepsilon}}_i\xrightarrow{p}\overline{\mathrm{E}\left[\boldsymbol{\varepsilon}_i'\boldsymbol{\Theta}_-'\mathbf{Q}_{T-1}'\ddot{\boldsymbol{\varepsilon}}_i\right]}=\overline{\mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{\Theta}_-'\mathbf{Q}_{T-1}'\ddot{\boldsymbol{\varepsilon}}_i\boldsymbol{\varepsilon}_i'\right)\right]}=\bar{\sigma}_\varepsilon^2\,\mathrm{tr}\left(\boldsymbol{\Theta}_-'\mathbf{Q}_{T-1}\mathbf{S}\right)=\bar{\sigma}_\varepsilon^2\,\mathrm{tr}\left(\mathbf{S}'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\right).$$

Finally, since α_i and $\boldsymbol{u}_i$, $\boldsymbol{u}_i$ and $\boldsymbol{\varepsilon}_i$, as well as α_i and $\boldsymbol{\varepsilon}_i$ are mutually independent, it immediately follows that

$$\frac{1}{N}\sum\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\xrightarrow{p}\left(\delta_Y+\tau\delta_X\right)^2\sigma_\alpha^2\boldsymbol{\iota}_T'\boldsymbol{\Theta}_-'\mathbf{Q}_{T-1}'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{\iota}_T+\left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)\mathrm{tr}\left(\boldsymbol{\Theta}_-'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\right).$$

To obtain the exact expressions, we note that

$$\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{\iota}_T=\mathbf{Q}_{T-1}\left(1,2,\ldots,T-1\right)'=\left(1-\frac{T}{2},2-\frac{T}{2},\ldots,(T-1)-\frac{T}{2}\right)'$$

such that

$$\begin{aligned}\boldsymbol{\iota}_T'\boldsymbol{\Theta}_-'\mathbf{Q}_{T-1}'\mathbf{Q}_{T-1}\boldsymbol{\Theta}_-\boldsymbol{\iota}_T &= \sum_{t=1}^{T-1}\left(t-\frac{T}{2}\right)^2=\sum_{t=1}^{T-1}t^2-2\frac{T}{2}\sum_{t=1}^{T-1}t+\sum_{t=1}^{T-1}\frac{T^2}{4}\\ &= \frac{(T-1)T(2T-1)}{6}-\frac{(T-1)T^2}{2}+\frac{(T-1)T^2}{4}\\ &= \frac{(T-1)T}{2}\left(\frac{2T-1}{3}-\frac{T}{2}\right)=\frac{T(T-1)(T-2)}{12}.\end{aligned}$$

Then,

$$\mathbf{Q}_{T-1}\mathbf{\Theta}_{-} = \begin{pmatrix} 0 & -\frac{T-2}{T-1} & -\frac{T-3}{T-1} & \cdots & \cdots & -\frac{1}{T-1} & 0 \\ 0 & \frac{1}{T-1} & -\frac{T-3}{T-1} & \cdots & \cdots & -\frac{1}{T-1} & 0 \\ 0 & \frac{1}{T-1} & \frac{2}{T-1} & \ddots & & -\frac{1}{T-1} & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \frac{1}{T-1} & \frac{2}{T-1} & & \ddots & -\frac{1}{T-1} & 0 \\ 0 & \frac{1}{T-1} & \frac{2}{T-1} & \cdots & \cdots & \frac{T-2}{T-1} & 0 \end{pmatrix}.$$

Since the ith row of $\mathbf{\Theta}'_{-}$ is the vector consisting of $T-1-i$ zeros and i ones, it is straightforward to compute the diagonal terms of $\mathbf{\Theta}'_{-}\mathbf{Q}_{T-1}\mathbf{\Theta}_{-}$ such that

$$\begin{aligned} \operatorname{tr}\left(\mathbf{\Theta}'_{-}\mathbf{Q}_{T-1}\mathbf{\Theta}_{-}\right) &= 0 + (T-2)\frac{1}{T-1} + (T-3)\frac{2}{T-1} + \ldots + 1\cdot\frac{T-2}{T-1} + 0 \\ &= \frac{1}{T-1}\sum_{t=1}^{T-2} t(T-1-t) = \frac{1}{T-1}\left((T-1)\sum_{t=1}^{T-2} t - \sum_{t=1}^{T-2} t^2\right) \\ &= \frac{(T-2)(4T-3)}{12}, \end{aligned}$$

and, analogously,

$$\operatorname{tr}\left(\mathbf{S}'\mathbf{Q}_{T-1}\mathbf{\Theta}_{-}\right) = 0 - \sum_{t=1}^{T-2}\frac{T-1-t}{T-1} + 0 = -\frac{T-2}{2}.$$

Summing up, this leads to

$$\begin{aligned} B_T^{FE-LDV} &= -\frac{\tau\bar{\sigma}_u^2\bar{\sigma}_\varepsilon^2\left(\operatorname{tr}\left(\mathbf{S}'\mathbf{Q}_{T-1}\mathbf{\Theta}_{-}\right)\right)^2}{(T-2)\,\bar{\sigma}_u^2\,\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-} - \tau^2\bar{\sigma}_u^4\left(\operatorname{tr}\left(\mathbf{S}'\mathbf{Q}_{T-1}\mathbf{\Theta}_{-}\right)\right)^2} \\ &= -\frac{\tau\bar{\sigma}_u^2\bar{\sigma}_\varepsilon^2\frac{(T-2)^2}{4}}{(T-2)\,\bar{\sigma}_u^2\left(\left(\delta_Y+\tau\delta_X\right)^2\sigma_\alpha^2\frac{T(T-1)(T-2)}{12} + \left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)\frac{(T-2)(4T-3)}{12}\right) - \tau^2\bar{\sigma}_u^4\frac{(T-2)^2}{4}} \\ &= -\tau\frac{\bar{\sigma}_\varepsilon^2}{\frac{1}{3}\left(\left(\delta_Y+\tau\delta_X\right)^2\sigma_\alpha^2(T-1)T + \left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)(4T-3)\right) - \tau^2\bar{\sigma}_u^2}. \end{aligned}$$

This will be smaller in magnitude than τ whenever

$$\frac{\bar{\sigma}_\varepsilon^2}{\frac{1}{3}\left(\left(\delta_Y+\tau\delta_X\right)^2\sigma_\alpha^2(T-1)T + \left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)(4T-3)\right) - \tau^2\bar{\sigma}_u^2} < 1,$$

or

$$\frac{1}{3}\left(\delta_Y+\tau\delta_X\right)^2\sigma_\alpha^2(T-1)T + \left(\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2\right)\left(\frac{4T-3}{3}-1\right) > 0,$$

which is the case for $T=3$ already.

B. Simulation Results

B.1. Robustness Tests

Table A1: Results when there is a feedback effect of $Y_{i,t-1}$ on X_{it} and a fixed effect simultaneously influences Y_{it} and X_{it}. Cases with small T and $\rho\tau$ close to unity.

	Sample size		Data generating process				Estimates of τ			
Scenario	N	T	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV
A	300	3	1.00	0.50	0.95	0.50	1.03	0.64	1.01	0.68
B	300	3	1.00	0.50	0.99	0.50	1.01	0.71	1.01	0.73

Results based on Monte Carlo Simulations with 500 repetitions. N: Number of individuals. T: Number of time periods

Table A2: Results when there is a feedback effect of $Y_{i,t-1}$ on X_{it} and a fixed effect simultaneously influences Y_{it} and X_{it}. Cases in which the FE-LDV estimate lies between the FE and LDV estimates (Scenario A) or all estimates are be downward biased (Scenario B).

	Sample size		Data generating process				Estimates of τ			
Scenario	N	T	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV
A	1000	6	1.00	-0.20	0.95	0.20	1.00	0.80	0.96	0.87
B	1000	3	1.00	-0.20	0.30	0.20	0.97	0.88	0.96	0.81

Results based on Monte Carlo Simulations with 500 repetitions. N: Number of individuals. T: Number of time periods

Table A3: Results when there is a feedback effect of $y_{i,t-1}$ on x_{it} and a fixed effect simultaneously influences y_{it} and x_{it}. Small ($\rho = 0.1$) vs high ($\rho = 0.9$) feedback effects.

	Data generating process				Estimates of τ			
Scenario	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV
A	1.00	0.50	0.10	0.50	1.22	0.98	1.13	0.98
B	1.00	-0.50	0.10	0.50	0.80	0.98	0.80	0.98
C	1.00	0.50	-0.10	0.50	1.17	1.02	1.15	0.98
D	1.00	-0.50	-0.10	0.50	0.80	1.02	0.80	0.99
E	-1.00	0.50	0.10	0.50	-0.80	-1.01	-0.80	-0.98
F	-1.00	-0.50	0.10	0.50	-1.17	-1.01	-1.15	-0.98
G	-1.00	0.50	-0.10	0.50	-0.80	-0.98	-0.80	-0.98
H	-1.00	-0.50	-0.10	0.50	-1.22	-0.98	-1.13	-0.97
I	1.00	0.50	0.90	0.50	1.05	0.80	1.02	0.88
J	1.00	-0.50	0.90	0.50	0.98	0.80	0.80	0.88
K	1.00	0.50	-0.90	0.50	1.00	1.01	1.20	1.00
L	1.00	-0.50	-0.90	0.50	0.98	1.01	0.80	1.00
M	-1.00	0.50	0.90	0.50	-0.97	-1.01	-0.80	-1.00
N	-1.00	-0.50	0.90	0.50	-1.00	-1.01	-1.20	-1.00
O	-1.00	0.50	-0.90	0.50	-0.98	-0.80	-0.80	-0.88
P	-1.00	-0.50	-0.90	0.50	-1.05	-0.80	-1.02	-0.88

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.

Table A4: Results when there is a feedback effect of $y_{i,t-1}$ on x_{it} and a fixed effect simultaneously influences y_{it} and x_{it}. The size of the feedback effect (ρ) depends on the fixed effect (α_i): $\rho = 0.5$ if $\alpha_i < 0$; $\rho = 0.1$ (in Scenarios A to D) or 0.9 (in Scenarios E to H) if $\alpha_i \geq 0$.

	Data generating process				Estimates of τ			
Scenario	τ	δ_X	ρ if $\alpha_i \geq 0$	δ_Y	OLS	FE	LDV	FE-LDV
A	1.00	0.50	0.10	0.50	1.21	0.93	1.10	0.95
B	1.00	-0.50	0.10	0.50	0.84	0.93	0.81	0.95
C	-1.00	0.50	0.10	0.50	-0.84	-1.02	-0.81	-0.98
D	-1.00	-0.50	0.10	0.50	-1.10	-1.03	-1.16	-0.98
E	1.00	0.50	0.90	0.50	1.07	0.84	1.01	0.91
F	1.00	-0.50	0.90	0.50	0.96	0.84	0.82	0.91
G	-1.00	0.50	0.90	0.50	-0.96	-1.01	-0.82	-0.98
H	-1.00	-0.50	0.90	0.50	-1.01	-1.01	-1.18	-0.98

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.
Idea behind this robustness test: In many applications, it may be the case that feedback effects are especially pronounced if an observation scores high on a specific fixed characteristic. For example, districts with strong partisan leanings might also show a stronger response to past election outcomes. To investigate the robustness of our findings in such a setting, we adapt the data generating process such that ρ is a function of $\alpha_i \sim N(0, 1)$.

Table A5: Results when there is a feedback effect of $y_{i,t-1}$ on x_{it} and a fixed effect simultaneously influences y_{it} and x_{it}. Differences in noise levels.

	Data generating process							Estimates of τ			
Scenario	τ	δ_X	ρ	δ_Y	σ_ϵ	σ_α	σ_u	OLS	FE	LDV	FE-LDV
A	1.00	0.50	0.50	0.50	5	1	1	1.06	0.53	1.18	0.89
B	1.00	-0.50	0.50	0.50	5	1	1	0.98	0.53	0.81	0.91
C	1.00	0.50	-0.50	0.50	5	1	1	1.01	1.13	1.20	0.99
D	1.00	-0.50	-0.50	0.50	5	1	1	0.97	1.14	0.80	0.98
E	-1.00	0.50	0.50	0.50	5	1	1	-0.98	-1.14	-0.80	-0.98
F	-1.00	-0.50	0.50	0.50	5	1	1	-1.01	-1.14	-1.20	-0.98
G	-1.00	0.50	-0.50	0.50	5	1	1	-0.97	-0.52	-0.80	-0.90
H	-1.00	-0.50	-0.50	0.50	5	1	1	-1.06	-0.53	-1.18	-0.90
I	1.00	0.50	0.50	0.50	1	5	1	1.32	0.89	1.14	0.95
J	1.00	-0.50	0.50	0.50	1	5	1	0.21	0.89	0.14	0.95
K	1.00	0.50	-0.50	0.50	1	5	1	1.88	1.03	1.55	0.99
L	1.00	-0.50	-0.50	0.50	1	5	1	0.21	1.03	1.14	0.99
M	-1.00	0.50	0.50	0.50	1	5	1	-0.21	-1.03	-1.14	-0.99
N	-1.00	-0.50	0.50	0.50	1	5	1	-1.88	-1.03	-1.55	-0.99
O	-1.00	0.50	-0.50	0.50	1	5	1	-0.21	-0.89	-0.14	-0.95
P	-1.00	-0.50	-0.50	0.50	1	5	1	-1.32	-0.89	-1.14	-0.95
Q	1.00	0.50	0.50	0.50	1	1	5	1.02	1.00	1.01	1.00
R	1.00	-0.50	0.50	0.50	1	1	5	0.99	0.99	0.99	1.00
S	1.00	0.50	-0.50	0.50	1	1	5	1.00	1.00	1.01	1.00
T	1.00	-0.50	-0.50	0.50	1	1	5	0.99	1.00	0.99	1.00
U	-1.00	0.50	0.50	0.50	1	1	5	-0.99	-1.00	-0.99	-1.00
V	-1.00	-0.50	0.50	0.50	1	1	5	-1.00	-1.00	-1.01	-1.00
W	-1.00	0.50	-0.50	0.50	1	1	5	-0.99	-0.99	-0.99	-1.00
X	-1.00	-0.50	-0.50	0.50	1	1	5	-1.02	-0.99	-1.01	-1.00

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.
Idea behind this robustness test: Some political variables may be measured much more noisily than others. Administrative data, for example, will often exhibit less noise than survey responses. Therefore, we systematically vary the standard deviations of the different disturbances in our data generating process.

Further robustness tests: The following listing provides an overview of several further scenarios we investigated to ensure that the FE-LDV model provides in absolute terms a lower bound of the true causal effect of X on Y. A check (✓) indicates that given the described change in the DGP and throughout all specifications presented in Tables 2, 3 and 6, the FE-LDV model provides an estimate that is in absolute terms lower than the true effect and has the correct sign.

- Change in the number of time periods from 6 to 3: ✓
- Change in the number of time periods from 6 to 16: ✓
- Change in the number of individuals from 300 to 10: ✓
- Change in the number of individuals from 300 to 1000: ✓
- Change in the value of δ_X and δ_Y from $\pm$ 0.5 to $\pm$ 1: ✓
- Change in the value of ρ from $\pm$ 0.5 to $\pm$ 0.95: ✓
- Change in the value of ρ from $\pm$ 0.5 to $\pm$ 0.05: ✓

B.2. State Dependence

Table A6: Monte Carlo Simulation results if the LDV model is correct, that is, $\rho \neq 0$ and there are no fixed effects: $\delta_X = 0 = \delta_Y$. Y_{it} follows a stationary AR(1) process (state dependence) according to Equation (8) with $\pi = 0.2$.

	Data generating process					Estimates of τ			
Scenario	τ	δ_X	ρ	δ_Y	π	OLS	FE	LDV	FE-LDV
A	1.00	0.00	0.50	0.00	0.20	1.20	0.97	1.00	0.92
B	1.00	0.00	-0.50	0.00	0.20	0.86	0.89	1.00	0.99
C	-1.00	0.00	0.50	0.00	0.20	-0.86	-0.89	-1.00	-0.99
D	-1.00	0.00	-0.50	0.00	0.20	-1.20	-0.97	-1.00	-0.92

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.

Table A7: Monte-Carlo Simulation results if the fixed effects model is correct, that is, $\delta_X \neq 0 \neq \delta_Y$ and there is no feedback effect: $\rho = 0$. Y_{it} follows a stationary AR(1) process (state dependence) according to Equation (8) with $\pi = 0.2$.

	Data generating process					Estimates of τ			
Scenario	τ	δ_X	ρ	δ_Y	π	OLS	FE	LDV	FE-LDV
E	1.00	0.50	0.00	0.50	0.20	1.30	0.96	1.12	0.97
F	1.00	-0.50	0.00	0.50	0.20	0.80	0.96	0.80	0.97
G	-1.00	0.50	0.00	0.50	0.20	-0.80	-0.96	-0.80	-0.97
H	-1.00	-0.50	0.00	0.50	0.20	-1.30	-0.96	-1.12	-0.97

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.

Table A8: Results when there is a feedback effect of $y_{i,t-1}$ on x_{it} and a fixed effect simultaneously influences y_{it} and x_{it}. Y_{it} follows a stationary AR(1) process (state dependence) according to Equation (8) with $\pi = 0.2$.

	Data generating process					Estimates of τ			
Scenario	τ	δ_X	ρ		π δ_Y	OLS	FE	LDV	FE-LDV
A	1.00	0.50	0.50	0.50	0.20	1.44	0.97	1.06	0.92
B	1.00	-0.50	0.50	0.50	0.20	1.06	0.97	0.80	0.92
C	1.00	0.50	-0.50	0.50	0.20	0.91	0.89	1.16	0.99
D	1.00	-0.50	-0.50	0.50	0.20	0.74	0.89	0.80	0.99
E	-1.00	0.50	0.50	0.50	0.20	-0.74	-0.89	-0.80	-0.99
F	-1.00	-0.50	0.50	0.50	0.20	-0.91	-0.89	-1.16	-0.99
G	-1.00	0.50	-0.50	0.50	0.20	-1.06	-0.97	-0.80	-0.92
H	-1.00	-0.50	-0.50	0.50	0.20	-1.44	-0.97	-1.06	-0.92

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.

B.3. Nonstationarity

Table A9: Monte Carlo Simulation results if the LDV model is correct, that is, $\rho \neq 0$ and there are no fixed effects: Unit root case ($\pi + \tau\rho = 1$).

	Data generating process				Estimates of τ			
Scenario	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV
A	1.00	0.00	1.00	0.00	1.00	0.79	1.00	0.86
B	1.00	0.00	-1.00	0.00	1.00	1.00	1.00	1.00
C	-1.00	0.00	1.00	0.00	-1.00	-1.00	-1.00	-1.00
D	-1.00	0.00	-1.00	0.00	-1.00	-0.78	-1.00	-0.86

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.

Table A10: Results when there is a feedback effect of $y_{i,t-1}$ on x_{it} and a fixed effect simultaneously influences y_{it} and x_{it}. Unit root case ($\pi + \tau\rho = 1$).

	Data generating process				Estimates of τ			
Scenario	τ	δ_X	ρ	δ_Y	OLS	FE	LDV	FE-LDV
A	1.00	0.50	1.00	0.50	1.01	0.91	1.01	0.94
B	1.00	-0.50	1.00	0.50	1.00	0.79	0.80	0.86
C	1.00	0.50	-1.00	0.50	1.00	1.00	1.20	1.00
D	1.00	-0.50	-1.00	0.50	1.00	1.00	0.80	1.00
E	-1.00	0.50	1.00	0.50	-1.00	-1.00	-0.80	-1.00
F	-1.00	-0.50	1.00	0.50	-1.00	-1.00	-1.20	-1.00
G	-1.00	0.50	-1.00	0.50	-1.00	-0.79	-0.80	-0.86
H	-1.00	-0.50	-1.00	0.50	-1.01	-0.91	-1.01	-0.94

Results based on Monte Carlo Simulations with 500 repetitions, 300 individuals and 6 time periods.

B.4. Empirical Example

Table A11: Replication of column (1) in Table 2 from Acemoglu et al., 2019 and estimation of the corresponding FE and LDV models

	(1) FE	(2) FE (restricted sample)	(3) LDV	(4) FE-LDV (replication of col. (1) in Table 2 from Acemoglu et al., 2019)
Democracy	-10.112	-9.815	0.457	0.973
	(4.316)	(4.375)	(0.296)	(0.294)
Log GDP, first lag			1.002	0.973
			(0.001)	(0.006)
Observations	6,934	6,790	6,790	6,790
Countries in sample	175	175	175	175

"This table presents estimates of the effect of democracy on the growth rate of GDP per capita. The reported coefficient on democracy is multiplied by 100. [...] In all speciations we control for a full set of country and year fixed effects. [...] Standard errors robust against heteroskedasticity and serial correlation at the country level are reported in parentheses." (Acemoglu et al., 2019, p. 67).
Since the static FE model allows the inclusion of an additional period in the estimation, the sample differs across columns. To rule out that the additional period is the cause of the strong difference between the FE model and the other models, we estimate an FE model in column (2) where the additional period is excluded, and the sample is therefore identical to that in columns (3) and (4).
These results are obtained by using the file "Table2_final.do" in the supplemental material "Data Archive" for Acemoglu et al. (2019, `https://www.journals.uchicago.edu/doi/suppl/10.1086/700936#`) and replacing the command in Line 463 by `xtreg y dem yy*, fe r cluster(wbcode2)` (column 1), `xtreg y dem yy* if l.y!=., fe r cluster(wbcode2)` (column 2), or `reg y l.y dem yy*, r cluster(wbcode2)` (column 3). After removing the prefix `quietly:`, the unchanged command line 463 (`xtreg y l.y dem yy*, fe r cluster(wbcode2)`) generates the results in column (4).

B.5. Simulations with varying T

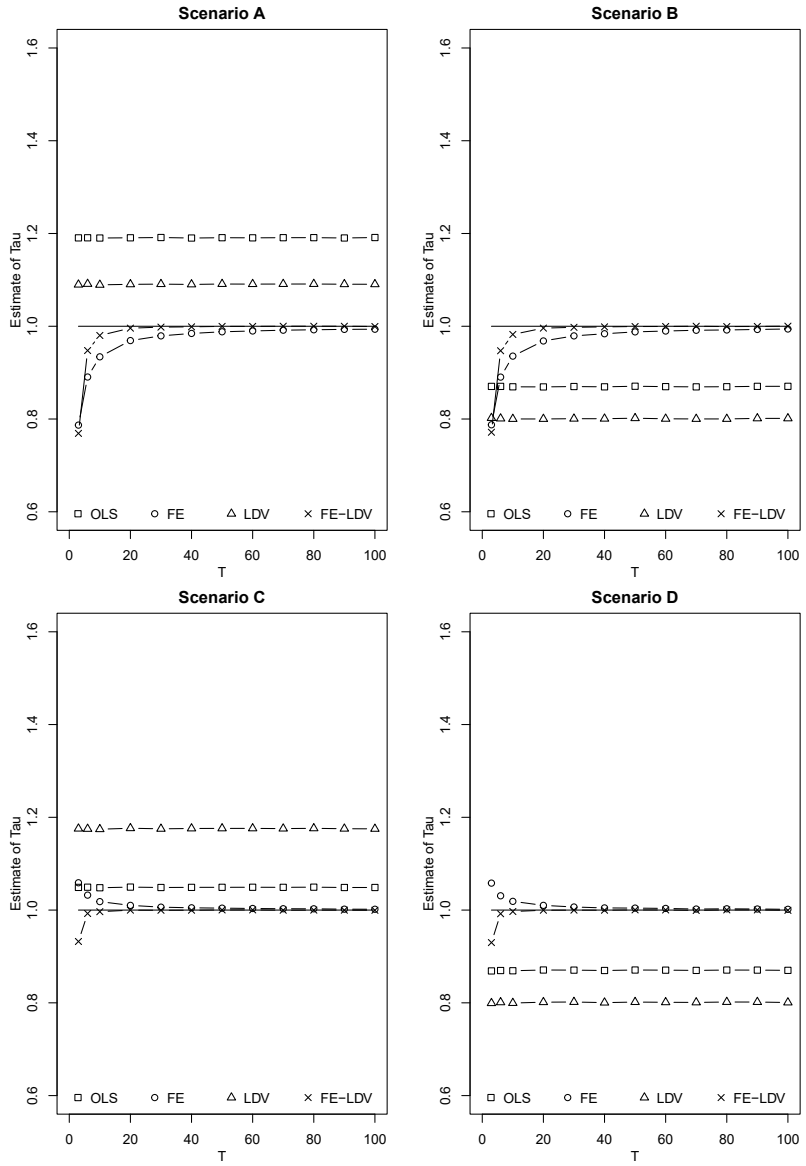

Figure A2: This figure extends the results shown in Table 6 of the main manuscript for different time periods T. With the exception of T, all other parameters are the same as in the corresponding scenarios in Table 6. Scenarios E to H are not shown here as they are exact mirror images of scenarios A to D in negative space.

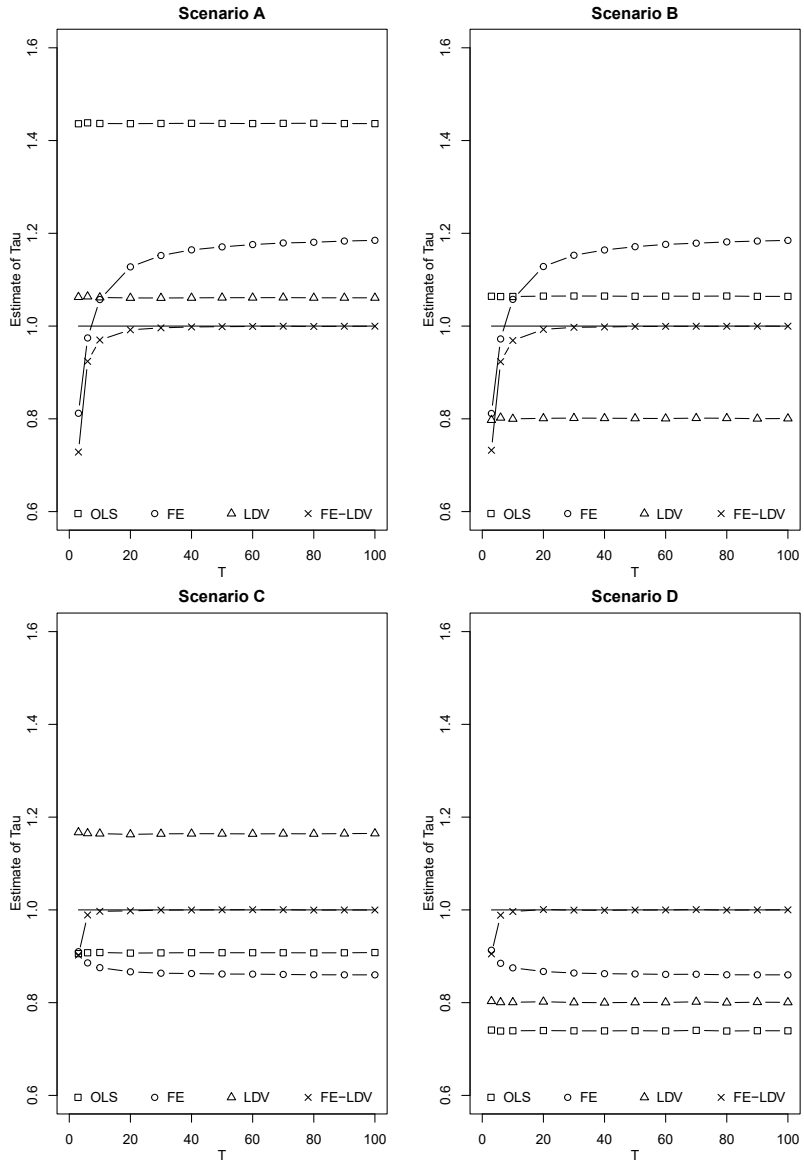

Figure A3: This figure extends the results shown in Table A8, the case of state dependence, for different time periods T. With the exception of T, all other parameters are the same as in the corresponding scenarios in Table A8. Scenarios E to H are not shown here as they are exact mirror images of scenarios A to D in negative space.

C. Technical Appendix

C.1. Preliminaries

For any $|\omega| < 1$, we approximate

$$\begin{aligned}\sum_{j=0}^{T-1} \omega^j &= \frac{1-\omega^T}{1-\omega} = \frac{1}{1-\omega} + O\left(e^{-T}\right) \\ \sum_{j=-T+1}^{T-1} \omega^{|j|} &= -1 + 2\sum_{j=0}^{T-1} \omega^j = \frac{1+\omega}{1-\omega} + O\left(e^{-T}\right),\end{aligned}$$

where the negligible terms vanish at exponential rates as T increases.[16] Furthermore,

$$\sum_{j=1}^{T-2} \frac{T-j}{T-1}\omega^j = \sum_{j=0}^{T-1} \omega^j - \frac{1}{T-1}\sum_{j=1}^{T-2} j\omega^j = \frac{1}{1-\omega} + O\left(e^{-T}\right) + O\left(T^{-1}\right)$$

since, with $|\omega| < 1$, the sum $\sum_{j=1}^{T-2} j\omega^j$ is easily shown to be bounded as T increases. Therefore, all three approximation errors above are of at most magnitude order $O\left(T^{-1}\right)$. Finally, for the same reasons,

$$\sum_{t=2}^{T}\sum_{s=2}^{T} \omega^{|s-t|} = (T-1)\frac{1+\omega}{1-\omega} + O\left(1\right).$$

C.2. Baseline DGP

C.2.1. Data generating process

Plugging in the regressor DGP in the panel model gives for $t = 1, \ldots, T$

$$\begin{aligned}y_{it} &= \delta_Y \alpha_i + \tau\left(\delta_X \alpha_i + \rho y_{i,t-1} + u_{it}\right) + \varepsilon_{it} \\ &= \left(\delta_Y + \tau\delta_X\right)\alpha_i + \tau\rho y_{i,t-1} + \tau u_{it} + \varepsilon_{it}.\end{aligned}$$

Let $\omega = \tau\rho$ and $v_{it} = \tau u_{it} + \varepsilon_{it}$.

Via the usual recursive argument, we immediately obtain for $t = 1, \ldots, T$

$$y_{it} = \omega^t y_{i0} + \left(\delta_Y + \tau\delta_X\right)\alpha_i \sum_{j=0}^{t-1} \omega^j + \sum_{j=0}^{t-1} \omega^j v_{i,t-j}.$$

Weak stationarity is then achieved by specifying the initial condition as

$$y_{i0} = \frac{\left(\delta_Y + \tau\delta_X\right)\alpha_i}{1-\omega} + \frac{v_{i0}}{\sqrt{1-\omega^2}},$$

[16] The famous expression for the Nickell bias was derived using similar approximations; see Nickell (1981).

where $|\tau\rho| = |\omega| < 1$, s.t. we have for $t = 1, \ldots, T$ that

$$y_{it} = \frac{(\delta_Y + \tau\delta_X)\,\alpha_i}{1-\omega} + \omega^t \frac{v_{i0}}{\sqrt{1-\omega^2}} + \sum_{j=0}^{t-1} \omega^j v_{i,t-j} \tag{A7}$$

and

$$y_{i,t-1} = \frac{(\delta_Y + \tau\delta_X)\,\alpha_i}{1-\omega} + \omega^{t-1} \frac{v_{i0}}{\sqrt{1-\omega^2}} + \sum_{j=0}^{t-2} \omega^j v_{i,t-1-j}, \tag{A8}$$

where the convention $\sum_i^j \cdot = 0$ for $j < i$ was used. Due to stationarity of y_{it} and the mutual independence of the error components, we have for $1 \leq s, t \leq T$

$$\mathrm{E}\left[y_{it} y_{is}\right] = \omega^{|t-s|} \frac{\tau^2\sigma_{ui}^2 + \sigma_{\varepsilon i}^2}{1-\omega^2} + \frac{(\delta_Y + \tau\delta_X)^2}{(1-\omega)^2}\left(\mu_\alpha^2 + \sigma_\alpha^2\right)$$

and

$$\mathrm{E}\left[\left(y_{it} - \frac{(\delta_Y + \tau\delta_X)\,\alpha_i}{1-\omega}\right)\left(y_{is} - \frac{(\delta_Y + \tau\delta_X)\,\alpha_i}{1-\omega}\right)\right] = \omega^{|t-s|}\frac{\tau^2\sigma_{ui}^2 + \sigma_{\varepsilon i}^2}{1-\omega^2}.$$

C.2.2. Within-transformed data

Since all variables are demeaned in a unit-specific manner that washes out any fixed components, we may set w.l.o.g. $\mu_\alpha, \sigma_\alpha^2 = 0$ in the following.

We first study the case of demeaning when an LDV is included as regressor (i.e., for the FE-LDV estimator). Naturally, this leads to an effectively shorter panel, $2 \leq t \leq T$. Then,

$$\begin{aligned}
\mathrm{E}\left[\ddot{y}_{it-1}^2\right] &= \mathrm{E}\left[y_{it-1}^2\right] - 2\,\mathrm{E}\left[\bar{y}_{i-} y_{it-1}\right] + \mathrm{E}\left[\bar{y}_{i-}^2\right] \\
&= \frac{\tau^2\sigma_{ui}^2 + \sigma_{\varepsilon i}^2}{1-\omega^2}\left(1 - \frac{2}{T-1}\sum_{j=2}^{T}\omega^{|t-j|} + \frac{1}{T-1}\left(1 + 2\sum_{j=1}^{T-2}\frac{T-j}{T-1}\omega^j\right)\right)
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right] &= \frac{\tau^2\sigma_{ui}^2 + \sigma_{\varepsilon i}^2}{1-\omega^2}\left(T - 1 - \frac{2}{T-1}\sum_{t=2}^{T}\sum_{j=2}^{T}\omega^{|t-j|} + \frac{1}{T-1}\sum_{t=2}^{T}\left(1 + 2\sum_{j=1}^{T-2}\frac{T-j}{T-1}\omega^j\right)\right) \\
&= (T-1)\,\frac{\tau^2\sigma_{ui}^2 + \sigma_{\varepsilon i}^2}{1-\omega^2}\left(1 - \frac{1}{T-1}\frac{1+\omega}{1-\omega}\right) + O\left(1\right)
\end{aligned}$$

for which we have as $N \to \infty$ that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right] &= (T-1)\,\frac{\tau^2\frac{1}{N}\sum_{i=1}^{N}\sigma_{ui}^2 + \frac{1}{N}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2}{1-\omega^2}\left(1 - \frac{1}{T-1}\frac{1+\omega}{1-\omega}\right) + O(1) \\
&\to (T-1)\,\frac{\tau^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{1-\omega^2}\left(1 - \frac{1}{T-1}\frac{1+\omega}{1-\omega}\right) + O(1),
\end{aligned}$$

since the $O(1)$ terms remain $O(1)$ upon averaging. For $T = 3$, the exact expression is

$$\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right] = \frac{\tau^2\sigma_{ui}^2 + \sigma_{\varepsilon i}^2}{1-\omega^2},$$

whose average converges as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\tau^2\sigma_{ui}^2 + \sigma_{\varepsilon i}^2}{1-\omega^2} = \frac{\tau^2\frac{1}{N}\sum_{i=1}^{N}\sigma_{ui}^2 + \frac{1}{N}\sum_{i=1}^{N}\sigma_{\varepsilon i}^2}{1-\omega^2} \to \frac{\tau^2\bar{\sigma}_u^2 + \bar{\sigma}_\varepsilon^2}{1-\omega^2}$$

Write then

$$\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{u}}_i\right] = \sum_{t=2}^{T}\mathrm{E}\left[\ddot{u}_{it}\ddot{y}_{it-1}\right].$$

Note now that $\mathrm{E}\left[u_{it}v_{is}\right] = \tau\sigma_{ui}^2$ for $s = t$ and 0 otherwise, such that

$$\mathrm{E}\left[u_{it}y_{is}\right] = \begin{cases}\tau\sigma_{ui}^2\omega^{s-t} & 2 \le t \le s < T \\ 0 & 1 \le s < t \le T\end{cases}.$$

This implies

$$\begin{aligned}\mathrm{E}\left[\ddot{u}_{it}\ddot{y}_{it-1}\right] &= \mathrm{E}\left[u_{it}y_{it-1}\right] - \mathrm{E}\left[\bar{u}_i y_{it-1}\right] - \mathrm{E}\left[\bar{y}_{i-}u_{it}\right] + \mathrm{E}\left[\bar{u}_i\bar{y}_{i-}\right] \\ &= \tau\sigma_{ui}^2\left(-\frac{1}{T-1}\sum_{j=0}^{t-3}\omega^j - \frac{1}{T-1}\sum_{j=0}^{T-1-t}\omega^j + \frac{1}{T-1}\sum_{j=0}^{T-3}\frac{T-j-2}{T-1}\omega^j\right),\end{aligned}$$

s.t.

$$\begin{aligned}\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{u}}_i\right] &= \tau\sigma_{ui}^2\left(-\frac{1}{T-1}\sum_{t=2}^{T}\sum_{j=0}^{t-3}\omega^j - \frac{1}{T-1}\sum_{t=2}^{T}\sum_{j=0}^{T-1-t}\omega^j + \frac{1}{T-1}\sum_{t=2}^{T}\sum_{j=0}^{T-3}\frac{T-j-2}{T-1}\omega^j\right) \\ &= (T-1)\,\tau\sigma_{ui}^2\left(-\frac{1}{T-1}\frac{1}{1-\omega}\right) + O(1);\end{aligned}$$

for $T = 3$, we obtain

$$\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{u}}_i\right] = -\frac{1}{2}\tau\sigma_{ui}^2,$$

and both quantities behave in the limit in an obvious manner upon averaging. Moreover,

$$\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right] = \sum_{t=2}^{T}\mathrm{E}\left[\ddot{\varepsilon}_{it}\ddot{y}_{it-1}\right],$$

where $\mathrm{E}\left[\varepsilon_{it}v_{is}\right] = \sigma_{\varepsilon i}^2$ for $s = t$ and 0 otherwise, such that

$$\mathrm{E}\left[\varepsilon_{it}y_{is}\right] = \begin{cases}\sigma_{\varepsilon i}^2\omega^{s-t} & 2 \le t \le s \le T \\ 0 & 1 \le s < t \le T\end{cases}$$

leading analogously to the case of $\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{u}}_i\right]$ to

$$\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right] = (T-1)\,\sigma_{\varepsilon i}^2\left(-\frac{1}{T-1}\frac{1}{1-\omega}\right) + O(1)$$

and

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]} = (T-1)\,\bar{\sigma}_\varepsilon^2\left(-\frac{1}{T-1}\frac{1}{1-\omega}\right) + O\left(1\right).$$

The exact expression is for $T = 3$

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{\varepsilon}}_i\right]} = -\frac{1}{2}\bar{\sigma}_\varepsilon^2.$$

We now move on to

$$\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\right] = \sum_{t=2}^{T}\mathrm{E}\left[\ddot{u}_{it}^2\right],$$

where

$$\begin{aligned}\mathrm{E}\left[\ddot{u}_{it}^2\right] &= \mathrm{E}\left[u_{it}^2\right] - 2\,\mathrm{E}\left[\bar{u}_i u_{it}\right] + \mathrm{E}\left[\bar{u}_{i-}^2\right] \\ &= \sigma_{ui}^2\left(1-\frac{1}{T-1}\right)\end{aligned}$$

such that

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\right]} = (T-1)\,\bar{\sigma}_u^2\left(1-\frac{1}{T-1}\right).$$

For $T = 3$, we have

$$\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\right] = \bar{\sigma}_u^2$$

(and note that there is no need for an approximation since $\mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\right]$ is tractable to begin with).

Finally,

$$\begin{aligned}\mathrm{E}\left[\ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{x}}_i\right] &= \rho^2\,\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{y}}_{i-}\right] + 2\rho\,\mathrm{E}\left[\ddot{\boldsymbol{y}}'_{i-}\ddot{\boldsymbol{u}}_i\right] + \mathrm{E}\left[\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{u}}_i\right] \\ &= (T-1)\left[\rho^2\frac{\tau^2\sigma_{ui}^2+\sigma_{\varepsilon i}^2}{1-\omega^2}\left(1-\frac{1}{T-1}\frac{1+\omega}{1-\omega}\right) + 2\rho\tau\sigma_{ui}^2\left(-\frac{1}{T-1}\frac{1}{1-\omega}\right)\right. \\ &\qquad \left. + \sigma_{ui}^2\left(1-\frac{1}{T-1}\right)\right] + O(1). \\ &= (T-1)\left(\rho^2\frac{\tau^2\sigma_{ui}^2+\sigma_{\varepsilon i}^2}{1-\omega^2} + \sigma_{ui}^2\right) + O(1),\end{aligned}$$

whose average converges,

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}'_i\ddot{\boldsymbol{x}}_i\right]} = (T-1)\left(\rho^2\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\omega^2} + \bar{\sigma}_u^2\right) + O(1),$$

Next we study the case when no LDV is included as regressor (i.e., for the simple FE estimator). The above expressions change slightly since, for $t = 1,\ldots,T$, $\ddot{x}_{it} = x_{it} - T^{-1}\sum_{t=1}^{T}x_{it}$. Using the same arguments as for the case with an LDV included, it is not difficult to obtain that

$$\overline{\mathrm{E}\left[\,\ddot{\boldsymbol{y}}'_{i-}\,\ddot{\boldsymbol{\varepsilon}}_i\right]} = T\bar{\sigma}_\varepsilon^2\left(-\frac{1}{T}\frac{1}{1-\omega}\right) + O\left(1\right)$$

and

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{x}}_i'\ddot{\boldsymbol{x}}_i\right]} = T\left(\rho^2\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\omega^2}+\bar{\sigma}_u^2\right)+O(1),$$

as well as

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{u}}_i\right]} = T\tau\bar{\sigma}_u^2\left(-\frac{1}{T}\frac{1}{1-\omega}\right)+O\left(1\right)$$

and

$$\overline{\mathrm{E}\left[\ddot{\boldsymbol{y}}_{i-}'\ddot{\boldsymbol{y}}_{i-}\right]} = T\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\omega^2}\left(1-\frac{1}{T}\frac{1+\omega}{1-\omega}\right)+O\left(1\right).$$

C.2.3. Pooled demeaning

Setting w.l.o.g. $\mu_\alpha = 0$, Equation (A8) together with the mutual independence of the error components implies immediately that

$$\mathrm{E}\left[\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right] = \sum_{t=2}^{T}\mathrm{E}\left[y_{it-1}^2\right] = (T-1)\left(\frac{\left(\delta_Y+\tau\delta_X\right)^2}{\left(1-\omega\right)^2}\sigma_\alpha^2+\frac{\tau^2\sigma_{ui}^2+\sigma_{\varepsilon i}^2}{1-\omega^2}\right).$$

Therefore, since $N^{-1}\sum_{i=1}^N\sigma_{ui}^2\to\bar{\sigma}_u^2$ and $N^{-1}\sum_{i=1}^N\sigma_{\varepsilon i}^2\to\bar{\sigma}_\varepsilon^2$,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[\boldsymbol{y}_{i-}'\boldsymbol{y}_{i-}\right] &= (T-1)\frac{\left(\delta_Y+\tau\delta_X\right)^2}{\left(1-\omega\right)^2}\sigma_\alpha^2+\frac{(T-1)}{N\left(1-\omega^2\right)}\sum_{i=1}^{N}\left(\tau^2\sigma_{ui}^2+\sigma_{\varepsilon i}^2\right)\\ &\to (T-1)\left(\frac{\left(\delta_Y+\tau\delta_X\right)^2}{\left(1-\omega\right)^2}\sigma_\alpha^2+\frac{\tau^2\bar{\sigma}_u^2+\bar{\sigma}_\varepsilon^2}{1-\omega^2}\right).\end{aligned}$$

Then, with $\boldsymbol{\delta}_X$ the column vector stacking δ_X $T-1$ times,

$$\begin{aligned}\mathrm{E}\left[\left(\boldsymbol{u}_i+\boldsymbol{\delta}_X\alpha_i\right)'\boldsymbol{y}_{i-}\right] &= \sum_{t=2}^{T}\mathrm{E}\left[y_{it-1}u_{it}\right]+\delta_X\sum_{t=2}^{T}\mathrm{E}\left[y_{it-1}\alpha_i\right]\\ &= (T-1)\frac{\delta_X\left(\delta_Y+\tau\delta_X\right)}{1-\omega}\sigma_\alpha^2\end{aligned}$$

not depending on i, and also obtain analogously

$$\mathrm{E}\left[\left(\boldsymbol{\varepsilon}_i+\boldsymbol{\delta}_Y\alpha_i\right)'\boldsymbol{y}_{i-}\right] = (T-1)\frac{\delta_Y\left(\delta_Y+\tau\delta_X\right)}{1-\omega}\sigma_\alpha^2.$$

Finally, we make use of the same arguments relying on the independence of the error components, leading immediately to

$$\begin{aligned}\mathrm{E}\left[\left(\boldsymbol{u}_i+\boldsymbol{\delta}_X\alpha_i\right)'\left(\boldsymbol{\varepsilon}_i+\boldsymbol{\delta}_Y\alpha_i\right)\right] &= \mathrm{E}\left[\boldsymbol{u}_i'\boldsymbol{\varepsilon}_i\right]+\mathrm{E}\left[\boldsymbol{u}_i'\boldsymbol{\delta}_Y\alpha_i\right]+\mathrm{E}\left[\boldsymbol{\delta}_X'\alpha_i\boldsymbol{\varepsilon}_i\right]+\mathrm{E}\left[\boldsymbol{\delta}_X'\boldsymbol{\delta}_Y\alpha_i^2\right]\\ &= (T-1)\,\delta_X\delta_Y\sigma_\alpha^2\end{aligned}$$

and

$$\mathrm{E}\left[\left(\boldsymbol{u}_i+\boldsymbol{\delta}_X\alpha_i\right)'\left(\boldsymbol{u}_i+\boldsymbol{\delta}_X\alpha_i\right)\right] = (T-1)\left(\sigma_{ui}^2+\delta_X^2\sigma_\alpha^2\right)$$

such that

$$\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[\left(\boldsymbol{u}_i+\boldsymbol{\delta}_X\alpha_i\right)'\left(\boldsymbol{u}_i+\boldsymbol{\delta}_X\alpha_i\right)\right]\to(T-1)\left(\bar{\sigma}_u^2+\delta_X^2\sigma_\alpha^2\right).$$

C.3. State-dependent DGP

We note that plugging in the regressor DGP in the panel model gives for $t=2,\ldots,T$

$$\begin{aligned} y_{it} &= \delta_Y\alpha_i+\pi y_{it-1}+\tau\left(\delta_X\alpha_i+\rho y_{it-1}+u_{it}\right)+\varepsilon_{it} \\ &= \left(\delta_Y+\tau\delta_X\right)\alpha_i+\left(\pi+\tau\rho\right)y_{it-1}+\tau u_{it}+\varepsilon_{it} \end{aligned}$$

and therefore can be nested in the original case by setting $\omega=\pi+\tau\rho$ instead of $\tau\rho$. Therefore all relevant expectations may be derived the same way as for the baseline setup, provided that the stationarity condition $|\pi+\tau\rho|<1$ is fulfilled.

References

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A Robinson (2019). "Democracy does cause growth". In: *Journal of Political Economy* 127.1, pp. 47–100.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arellano, Manuel and Stephen Bond (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". In: *The Review of Economic Studies* 58.2, pp. 277–297.

Arkhangelsky, Dmitry and Guido Imbens (2024). "Causal models for longitudinal and panel data: A survey". In: *The Econometrics Journal* 27.3, pp. C1–C61.

Beck, Nathaniel and Jonathan N Katz (2011). "Modeling dynamics in time-series–cross-section political economy data". In: *Annual Review of Political Science* 14.1, pp. 331–352.

Beckmann, Michael and Matthias Kräkel (2022). "Empowerment, task commitment, and performance pay". In: *Journal of Labor Economics* 40.4, pp. 889–938.

Bun, Maurice JG and Frank Windmeijer (2010). "The weak instrument problem of the system GMM estimator in dynamic panel data models". In: *The Econometrics Journal* 13.1, pp. 95–126.

Chow, Y. S. (1971). "On the L_p-Convergence for $n^{-1/p}S_n$, $0 < p < 2$". In: *The Annals of Mathematical Statistics* 42.1, pp. 393–394.

Chudik, Alexander and M Hashem Pesaran (2019). "Mean group estimation in presence of weakly cross-correlated estimators". In: *Economics Letters* 175, pp. 101–105.

Dee, Laura E, Paul J Ferraro, Christopher N Severen, Kaitlin A Kimmel, Elizabeth T Borer, Jarrett EK Byrnes, Adam Thomas Clark, Yann Hautier, Andrew Hector, Xavier Raynaud, et al. (2023). "Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and methods for causal inference". In: *Nature Communications* 14.1, p. 2607.

Ding, Peng and Fan Li (2019). "A bracketing relationship between difference-in-differences and lagged-dependent-variable adjustment". In: *Political Analysis* 27.4, pp. 605–615.

Falk, Armin, Fabian Kosse, Ingo Menrath, Pablo E Verde, and Johannes Siegrist (2018). "Unfair pay and health". In: *Management Science* 64.4, pp. 1477–1488.

Frondel, Manuel and Colin Vance (2010). "Fixed, random, or something in between? A variant of Hausman's specification test for panel data estimators". In: *Economics Letters* 107.3, pp. 327–329.

Guryan, Jonathan (2001). *Desegregation and Black Dropout Rates*. NBER Working Paper No. 8345.

Hausman, Jerry A (1978). "Specification tests in econometrics". In: *Econometrica*, pp. 1251–1271.

Hsiao, Cheng (2022). *Analysis of Panel Data*. fourth. Cambridge University Press.

Kampkötter, Patrick and Dirk Sliwka (2018). "More dispersion, higher bonuses? On differentiation in subjective performance evaluations". In: *Journal of Labor Economics* 36.2, pp. 511–549.

Keele, Luke, William Cubbison, and Ismail White (2021). "Suppressing Black votes: A historical case study of voting restrictions in Louisiana". In: *American Political Science Review* 115.2, pp. 694–700.

Keele, Luke and Nathan J Kelly (2006). "Dynamic models for dynamic theories: The ins and outs of lagged dependent variables". In: *Political Analysis* 14.2, pp. 186–205.

Keele, Luke, Neil Malhotra, and Colin H McCubbins (2013). "Do term limits restrain state fiscal policy? Approaches for causal inference in assessing the effects of legislative institutions". In: *Legislative Studies Quarterly* 38.3, pp. 291–326.

Klosin, Sylvia (2024). *Dynamic Biases of Static Panel Data Estimators*. arXiv: 2410.16112 [econ.EM]. URL: https://arxiv.org/abs/2410.16112.

Leszczensky, Lars and Tobias Wolbring (2022). "How to deal with reverse causality using panel data? Recommendations for researchers based on a simulation study". In: *Sociological Methods & Research* 51.2, pp. 837–865.

Marsh, Wayde ZC (2023). "Trauma and turnout: The political consequences of traumatic events". In: *American Political Science Review* 117.3, pp. 1036–1052.

Moral-Benito, Enrique (2013). "Likelihood-based estimation of dynamic panels with predetermined regressors". In: *Journal of Business & Economic Statistics* 31.4, pp. 451–472.

Moral-Benito, Enrique, Paul Allison, and Richard Williams (2019). "Dynamic panel data modelling using maximum likelihood: An alternative to Arellano-Bond". In: *Applied Economics* 51.20, pp. 2221–2232.

Nickell, Stephen (1981). "Biases in Dynamic Models with Fixed Effects". In: *Econometrica* 49.6, pp. 1417–1426.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe (2023). "What's trending in difference-in-differences? A synthesis of the recent econometrics literature". In: *Journal of Econometrics* 235.2, pp. 2218–2244.

Tomberg, Lukas, Karen Smith Stegen, and Colin Vance (2021). ""The mother of all political problems"? On asylum seekers and elections". In: *European Journal of Political Economy* 67, p. 101981.

von Hippel, Paul T (2022). "Is psychological science self-correcting? Citations before and after successful and failed replications". In: *Perspectives on Psychological Science* 17.6, pp. 1556–1565.

Wooldridge, Jeffrey M (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.